

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:32:13 ; Search time 30 seconds
(without alignments)
846.666 Million cell updates/sec

Title: US-09-615-285B-2

Perfect score: 2717
Sequence: 1 MALNSGSPPAIGPYENHGY.....YGVNVVFTDVIYQMRADG 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pap:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pap:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pap:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pap:*
5: /cgn2_6/prodata/2/iaa/PCUTUS_COMB.pap:*
6: /cgn2_6/prodata/2/iaa/backfiles.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2717	100.0	492	3	US-09-342-749-2
2	2717	100.0	492	4	US-09-691-840-2
3	2696	99.2	492	4	US-09-685-166A-895
4	1540	56.7	283	3	US-08-807-151-1
5	1540	56.7	283	4	US-09-478-957-1
6	1165	42.9	209	4	US-09-685-166A-897
7	898	32.7	454	3	US-09-518-046-2
8	856	31.9	159	3	US-09-518-046-24
9	780.5	28.7	455	3	US-09-261-416-2
10	694	25.2	423	4	US-09-656-002-2
11	676.5	24.9	406	4	US-09-851-588-6
12	676.5	24.9	435	3	US-09-008-271A-6
13	676.5	24.9	437	4	US-09-851-588-8
14	660	24.3	798	1	US-08-200-900A-2
15	660	24.3	798	5	PCT-US94-00616-2
16	565.5	20.8	417	4	US-08-820-002-4
17	558.5	20.6	855	2	US-09-027-337-2
18	558.5	20.6	855	4	US-09-644-600-2
19	558.5	20.6	855	4	US-09-654-600A-2
20	557	20.5	376	4	US-09-820-002-2
21	553.5	20.4	638	2	US-08-944-483-63
22	548.5	20.2	248	3	US-08-681-151-3
23	548.5	20.2	902	4	US-09-644-600-10
24	548.5	20.2	902	4	US-09-654-600A-10
25	539.5	19.9	416	2	US-08-000-846-2
26	534	19.7	356	1	US-08-681-151-1
27	528.5	19.5	256	2	US-09-027-337-3

28	528.5	19.5	256	4	US-09-644-600-3	Sequence 3, Appli
29	528.5	19.5	256	4	US-09-654-600A-3	Sequence 3, Appli
30	527.5	19.4	418	1	US-08-508-448C-25	Sequence 25, Appli
31	527.5	19.4	418	4	US-09-370-838-62	Sequence 62, Appli
32	527.5	19.4	418	4	US-09-370-838-82	Sequence 82, Appli
33	527.5	19.4	418	4	US-09-370-838-83	Sequence 83, Appli
34	524	19.3	238	3	US-08-944-483-64	Sequence 64, Appli
35	523.5	19.3	255	3	US-08-944-483-67	Sequence 67, Appli
36	521.5	19.2	235	3	US-08-807-151-3	Sequence 3, Appli
37	521.5	19.2	235	4	US-09-478-357-3	Sequence 3, Appli
38	518.5	19.1	812	1	US-08-248-629A-1	Sequence 1, Appli
39	518.5	19.1	812	1	US-08-451-932-1	Sequence 1, Appli
40	518.5	19.1	812	1	US-08-452-360-1	Sequence 1, Appli
41	518.5	19.1	812	1	US-08-326-785-1	Sequence 1, Appli
42	518.5	19.1	812	2	US-08-612-788-1	Sequence 1, Appli
43	518.5	19.1	812	2	US-08-605-598B-1	Sequence 1, Appli
44	518.5	19.1	812	2	US-08-429-743-1	Sequence 1, Appli
45	518.5	19.1	812	2	US-08-866-735-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: Tumor suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match	100.0%;	Score	2717;	DB	3;	Length	492;
Best Local Similarity	100.0%;	Pred. No.	5.3e-244;				
Matches	492;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MALNSGSPPAIGPYENHGYQENPYPAQPTVVPTVYVHPAQYVPSVPVQVAPRVLTOA	60				
Db	1	MALNSGSPPAIGPYENHGYQENPYPAQPTVVPTVYVHPAQYVPSVPVQVAPRVLTOA	60				
QY	61	SNPVCTQPKSPSGTCTCTKTCALCITITLGTFLVGAALAGLLKFMGSKCSNGIEC	120				
Db	61	SNPVCTQPKSPSGTCTCTKTCALCITITLGTFLVGAALAGLLKFMGSKCSNGIEC	120				
QY	121	DSGTCINPSNWCDSVSHCPGGDENRCVRLYGPNFILQVYSQKSWHPVQDDWNNY	180				
Db	121	DSGTCINPSNWCDSVSHCPGGDENRCVRLYGPNFILQVYSQKSWHPVQDDWNNY	180				
QY	181	GRAACRDMGYKNFYSSQGIIVDDSGTSPMKLNTSAGNVDIYKLYHSDACSSKAVSLR	240				
Db	181	GRAACRDMGYKNFYSSQGIIVDDSGTSPMKLNTSAGNVDIYKLYHSDACSSKAVSLR	240				
QY	241	CIACGVNLNSRQSRIVGGESALPGAWPQVSLHVNHVHVCSSIIITPEWITTAHCVEK	300				
Db	241	CIACGVNLNSRQSRIVGGESALPGAWPQVSLHVNHVHVCSSIIITPEWITTAHCVEK	300				
QY	301	PLANPHTWTAFAQILRQSFYAGVQVEKVTSHNYSKTKKNDIALMKLQKPLTFNDL	360				
Db	301	PLANPHTWTAFAQILRQSFYAGVQVEKVTSHNYSKTKKNDIALMKLQKPLTFNDL	360				

QY 361 VKPVCLEPNPQMLQPEQLCWSGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480
DB 361 VAPVCLPFPQMLQPEQLCWSGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 420
QY 421 TPAMICAGFLOGNVDSQCGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480
DB 421 TPAMICAGFLOGNVDSQCGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480
QY 481 TDWIYRQMRADG 492
DB 481 TDWIYRQMRADG 492

RESULT 2
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-691-840-2

Query Match 100.0%; Score 2717; DB 4; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.3e-244;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGPPAIGPYENHGYQENPYPAQPTVTVYVHVAQYYPSPVQVAPRVLTOA 60
DB 1 MALNSGPPAIGPYENHGYQENPYPAQPTVTVYVHVAQYYPSPVQVAPRVLTOA 60
QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWKFGKCSNSGIEC 120
DB 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWKFGKCSNSGIEC 120
QY 121 DSSGTCINPNCWCDGVSHCPGGEDENRVLYGPNFLOVYSSQKSWHPVCQDDWNNY 180
DB 121 DSSGTCINPNCWCDGVSHCPGGEDENRVLYGPNFLOVYSSQKSWHPVCQDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIRVGGESALPGAWPQVSLHVQNVHVCVGSIIITPEWIVTAACHVEK 300
DB 241 CIACGVNLNSRQSRIRVGGESALPGAWPQVSLHVQNVHVCVGSIIITPEWIVTAACHVEK 300
QY 301 PLANNPFWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTNNDIALMKLQKPLTFNDL 360
DB 301 PLANNPFWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTNNDIALMKLQKPLTFNDL 360
QY 361 VKPVCLEPNPQMLQPEQLCWSGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 420
DB 361 VKPVCLEPNPQMLQPEQLCWSGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 420
QY 421 TPAMICAGFLOGNVDSQCGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480

DB 421 TPAMICAGFLOGNVDSQCGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480
QY 481 TDWIYRQMRADG 492
DB 481 TDWIYRQMRADG 492

RESULT 3
US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895

Query Match 99.2%; Score 2696; DB 4; Length 492;
Best Local Similarity 98.8%; Pred. No. 4.7e-242;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGPPAIGPYENHGYQENPYPAQPTVTVYVHVAQYYPSPVQVAPRVLTOA 60
DB 1 MALNSGPPAIGPYENHGYQENPYPAQPTVTVYVHVAQYYPSPVQVAPRVLTOA 60
QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWKFGKCSNSGIEC 120
DB 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWKFGKCSNSGIEC 120
QY 121 DSSGTCINPNCWCDGVSHCPGGEDENRVLYGPNFLOVYSSQKSWHPVCQDDWNNY 180
DB 121 DSSGTCINPNCWCDGVSHCPGGEDENRVLYGPNFLOVYSSQKSWHPVCQDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDMGYKNFYSSQGIYVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIRVGGESALPGAWPQVSLHVQNVHVCVGSIIITPEWIVTAACHVEK 300
DB 241 CIACGVNLNSRQSRIRVGGESALPGAWPQVSLHVQNVHVCVGSIIITPEWIVTAACHVEK 300
QY 301 PLANNPFWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTNNDIALMKLQKPLTFNDL 360
DB 301 PLANNPFWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTNNDIALMKLQKPLTFNDL 360
QY 361 VKPVCLEPNPQMLQPEQLCWSGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 420
DB 361 VKPVCLEPNPQMLQPEQLCWSGSGGGLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 420

QY 421 TPAMICAGFLQGNVDSQDGGPLVTSKNNIWMWLGDTSMGSCCAKAYRQVYGNVNF 480
Db 421 TPAMICAGFLQGNVDSQDGGPLVTSNNIWMWLGDTSMGSCCAKAYRQVYGNVNF 480
QY 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRANG 492

RESULT 4
US-08-807-151-1
; Sequence 1, Application US/08807151
; Patent No. 6043033
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNT01
; CLONE: 556016
US-08-807-151-1

Query Match 56.7%; Score 1540; DB 3; Length 283;
Best Local Similarity 99.6%; Pred. No. 6.1e-135;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 210 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCLACGVNLSRSQSRIVGGESALPGAWPW 269
Db 1 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCLACGVNLSRSQSRIVGGESALPGAWPW 60
QY 270 QVSLHVQNVHVCSSIIITPEWIVTAAHCVKELNPNPWHWTAFAGLRQSFMYGAGYQVE 329
Db 61 QVSLHVQNVHVCSSIIITPEWIVTAAHCVKELNPNPWHWTAFAGLRQSFMYGAGYQVE 120
QY 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVPCLNPGMTLQPEOLCWSISWGATEE 389
Db 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVPCLNPGMTLQPEOLCWSISWGATEE 180

QY 390 KGKTSVLNAAKVLLIETORCNSRVVYDNLITPAMICAGFLQGNVDSQDGGPLVTSK 449
Db 181 KGKTSVLNAAKVLLIETORCNSRVVYDNLITPAMICAGFLQGNVDSQDGGPLVTSK 240
QY 450 NNIMWLGDTSMGSCCAKAYRQVYGNVNFVTDWIYRQMRADG 492
Db 241 NNIMWLGDTSMGSCCAKAYRQVYGNVNFVTDWIYRQMRADG 283
RESULT 5
US-09-478-957-1
; Sequence 1, Application US/09478957
; Patent No. 6350448
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/478,957
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,151
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNT01
; CLONE: 556016
US-09-478-957-1

Query Match 56.7%; Score 1540; DB 4; Length 283;
Best Local Similarity 99.6%; Pred. No. 6.1e-135;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 210 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCLACGVNLSRSQSRIVGGESALPGAWPW 269
Db 1 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCLACGVNLSRSQSRIVGGESALPGAWPW 60
QY 270 QVSLHVQNVHVCSSIIITPEWIVTAAHCVKELNPNPWHWTAFAGLRQSFMYGAGYQVE 329
Db 61 QVSLHVQNVHVCSSIIITPEWIVTAAHCVKELNPNPWHWTAFAGLRQSFMYGAGYQVE 120
QY 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVPCLNPGMTLQPEOLCWSISWGATEE 389
Db 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVPCLNPGMTLQPEOLCWSISWGATEE 180
QY 390 KGKTSVLNAAKVLLIETORCNSRVVYDNLITPAMICAGFLQGNVDSQDGGPLVTSK 449

Db 181 KGTSEVLNAKVLLETORCNSRVVYDNLITPAMICAGFLQGNVDSGGXLTWSK 240
Qy 450 NNIMWLIGTSGSCAKAYRPGVYGNVMTDVIYROMRADG 492
Db 241 NNIMWLIGTSGSCAKAYRPGVYGNVMTDVIYROMRADG 283

RESULT 6

US-09-685-166A-897
; Sequence 897, Application US/09685166A
; Patent No. 630305
; GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedwick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

FILE REFERENCE: 210121.427C21

CURRENT APPLICATION NUMBER: US/09/685,166A

CURRENT FILING DATE: 2000-10-10

NUMBER OF SEQ ID NOS: 898

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 897

LENGTH: 209

TYPE: PRT

ORGANISM: Homo sapiens

US-09-685-166A-897

Query Match 42.9%; Score 1165; DB 4; Length 209;
Best Local Similarity 99.5%; Pred. No. 2.8e-100; Indels 0; Gaps 0;
Matches 208; Conservative 1; Mismatches 0;

Qy 1 MALNSGPPAIGPYENHGYQENPYPAQPTVPTVYVHPAQYPPSPVQYAPRVLQA 60
Db 1 MALNSGPPAIGPYENHGYQENPYPAQPTVPTVYVHPAQYPPSPVQYAPRVLQA 60
Qy 61 SNPVTCTPKSPGVTCTSKKALCITLTGTLVGAALAGLLKFMGSKCSNSGIEC 120
Db 61 SNPVTCTPKSPGVTCTSKKALCITLTGTLVGAALAGLLKFMGSKCSNSGIEC 120
Qy 121 DSSGTCINPSNWCDCVSHCPGDEENRCVRLYGNPFIQVYSSQKSWHPVCDDWNNY 180
Db 121 DSSGTCINPSNWCDCVSHCPGDEENRCVRLYGNPFIQVYSSQKSWHPVCDDWNNY 180
Qy 181 GRAACRDMGYKNFYSSGIVDDSGSTSF 209
Db 181 GRAACRDMGYKNFYSSGIVDDSGSTSF 209

RESULT 7

US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

APPLICANT: Underwood, Lowell J.

TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed

TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof

FILE REFERENCE: D6192CIP
CURRENT APPLICATION NUMBER: US/09/518,046
CURRENT FILING DATE: 2000-03-02
EARLIER APPLICATION NUMBER: 09/261,416
EARLIER FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 153
SEQ ID NO 2
LENGTH: 454
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: complete amino acid sequence of TADG-12
OTHER INFORMATION: protein
US-09-518-046-2

Query Match 32.7%; Score 888; DB 3; Length 454;

Best Local Similarity 46.3%; Pred. No. 4.9e-74;

Matches 190; Conservative 52; Mismatches 148; Indels 20; Gaps 10;

Qy 90 TLGTLVGAALAGLLKFMGSKCSNSGIECDSTGTCINPSNWCDCVSHCPGDEENRCV 149
Db 54 SLGIITLILALAILGLIHF---DCSGK-YRCRSFKCIELITRCDGVSDCKDGEYRCV 109
Qy 150 RLYGFNFILQVYSSQKSWHPVCDDWNNYGRACRDMGYKNFYSSQKSWHPVCDDWNNY 209
Db 110 RVGGONAVLQVFTA--ASWKTMCSDDKWGHYANVACAQLGFP-SYVSSDNLRYVSLGQF 166
Qy 210 MKLNTSAGNV---DIYKKLYHS---DACSSKAVSLRCIACGVNLNSRQSRIVGESA 262
Db 187 REEFVSDIHLDPDKVTALHSHSVYREGCASGHVILQCTACGHRGYS--SRIVGNMS 224
Qy 263 LPGAHPQVSLHVNQVHVCVGSITPEWITAAHCYKPLNNPWHMTAFAGILRQSFMY 322
Db 225 LLSQWPQASLQFGYHLCGSSVITPLWITAAHCV-YDLYLPKSWTIQVGLV--SLLDN 281
Qy 323 GA-GYQVEKVI SHENYDSKTKNDIALMKLQKLTENDLVKPYCLNPGMWLQPEQLCWI 381
Db 282 PAPSHLVEKVIYHSHYKPKRLGNDIALMKLAGLPTFNEMIQPVCLNSEENFPDGVKVCWT 341
Qy 382 SGWGATEEKGKTSVNLNAKVLLETORCNSRVVYDNLITPAMICAGFLQGNVDSGGX 441
Db 342 SGWGATEDGGDASPLNHAAPVPLISKNICNHRDVGGIISPSMLCAGYLTGGVDSQGD 401
Qy 442 GGPLVTSKNINWLIGTSGSCAKAYRPGVYGNVMTDVIYROMRAD 491
Db 402 GGPLVTCERRLKLKLVGATSGFAGCAEVNKGIVTRVTSFLDNIHQMERD 451

RESULT 8

US-09-518-046-24
; Sequence 24, Application US/09518046
; Patent No. 6294663

GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

APPLICANT: Underwood, Lowell J.

TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed

TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof

FILE REFERENCE: D6192CIP

CURRENT APPLICATION NUMBER: US/09/518,046

CURRENT FILING DATE: 2000-03-02

EARLIER APPLICATION NUMBER: 09/261,416

EARLIER FILING DATE: 1999-03-03

NUMBER OF SEQ ID NOS: 153

SEQ ID NO 24

LENGTH: 159

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: DOMAIN

OTHER INFORMATION: protease domain of TMRPS2 (Tmrps2)

US-09-518-046-24

Query Match 31.9%; Score 866; DB 3; Length 159;

Best Local Similarity 99.4%; Pred. No. 1.2e-72;
Matches 158; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 289 EHWTAACHVCKPLNNPHWHTAFAGILRQSMFYGAGYQVEKVISHPNVDSTKKNNDIAL 348
Db 1 EHWVTAACHVCKPLNNPHWHTAFAGILRQSMFYGAGYQVEKVISHPNVDSTKKNNDIAL 60
Qy 349 MKLQKPLFNDLVKPVCLPNNPMMQLQPEQLCWISGAGTEKGTSEVLNAKAVLLIETQ 408
Db 61 MKLQKPLFNDLVKPVCLPNNPMMQLQPEQLCWISGAGTEKGTSEVLNAKAVLLIETQ 120
Qy 409 RNSRYVVDNLIITPAMICAGFLOGNVDSGCGSGPLVT 447
Db 121 RNSRYVVDNLIITPAMICAGFLOGNVDSGCGSGPLVT 159

RESULT 9

US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261.416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TAGD-12 encoded by nucleotides
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
US-09-261-416-2

Query Match 25.7%; Score 780.5; DB 3; Length 455;

Best Local Similarity 43.6%; Pred. No. 4.8e-64;
Matches 180; Conservative 50; Mismatches 158; Indels 25; Gaps 13;
Qy 90 TLGFLVGAALAAAGLLMKFGSKNSGIECDSSGTCINPSNWCDDGVSHCPGGEDEKCV 149
Db 54 SLGIALLALALGLGHF--DCSGK-YRCRSFKIELITRCDGVSDCKGDEBYRCV 109
Qy 150 RLYGFNFLOYSSQRKSHFVCOODNENYGRAACRDMGYKXNFYSSQGIYDDSGSTSP 209
Db 110 RVGGQNAVLOVFTA--ASWTKMGSDDWKGYANVACAQLGFP--SYVSSDNLRVSLGQF 166
Qy 210 MKLNTSAGNV---DIYKLYHS---DACS KAVWSIRCIACGVNLSRSRQRIYGGESA 262
Db 167 REEFVSDHLLPDDKVTAHSHVYVREGCASGHVVTLOCTACGHRGYS--SRIYVGNMS 224
Qy 263 LPGAWPQVSLHVQNVHVGCGSIITPEWIVTAACHVCKEPLNNPHWHTAFAGILRQSMFY 322
Db 225 LLSQMPWQASLQFQYHLCGSGVITPLWITAAHCY-YDLYLKSWTIQVGLV--SLLDN 281
Qy 323 GA-GYQVEKVISHPNVDSTKKNNDIALMKLQKPLTENDLVKPVCLNPGMMLQPOLCWI 381
Db 282 PAPSHLVKIVYHSKYKPKLGNNDIALMKLAGELTNEMIQVCLPNSSEENPPDKVCWT 341
Qy 382 SGWAGTEKGTSEVLNAKAVLLIETQ--RNSRYVVDNLIITPAMICAGFLOG-NVDSQ 438
Db 342 SGWAGTEGDDGASVPLNHAAPVLIISNKDLPQQRVWRHHL--FLHALRGLPDGWRWNSCQ 399
Qy 439 GDSGGLVTSKNNIWLIGTSGWGCACAYRPGVYGNVWFTDVIYQMRAD 491
Db 400 GDSGGLVQCERRLKLVGATSGIGCADVYKGVVTRVTSFLDVIHQMERD 452

RESULT 10

US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; FILE REFERENCE: A-69108/DJ8/JUD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-656-002-2

Query Match 25.2%; Score 684; DB 4; Length 423;

Best Local Similarity 36.2%; Pred. No. 4e-35;
Matches 165; Conservative 64; Mismatches 167; Indels 60; Gaps 16;
Qy 61 SNPVCTQPKSPGTVCTKTKKALCITITLGTFLVGAALAAAGLLMKFGSKNSGIEC 120
Db 2 SNP--CANPVSP-WRPESVGIPLIALLSLASIIIVVLLIKVILDKYY-----FLC 50
Qy 121 DSSGTCINPSNWCDDGVSHCPGGEDEKCVRLY--GP-----NFTLOYVSSQRKSH 169
Db 51 GQPLHFIPKQCLDGLDCLPGLGEDEHCVKSPFPGPAVAVLRSKDRSTLOVLDATGNWF 110
Qy 170 PVCQDDNENYGRAACRDMGY--KXNFYSSQ-----GIVDDSGSTSFMKLNTSAGNVD 220
Db 111 SADFONFTEALAEATACRQMGYSKPTFRAVIGPDQLDVVVEITENSCELMRNSSG--- 167
Qy 221 IYKLVHSDACSKKAVWSIRCIACGVNLSRSRQRIYGGESALPGAWPQVSLHVQNVHV 280
Db 168 -----PCLSGSELVHLCLACGSL---KTRFVVGGEASVDSWFWQVSIQYDKQHV 215
Qy 281 CGSIIITPEWIVTAACHVCKEPLNNPHWHTAFAGILR-QSFMFYGAGYQVEKVI---SHPN 336
Db 216 CGSILDPHWLTAACHCFRKH-TDVFNNKVRAGSDKLGSP-----PSLAVAKIIIEFNPM 270
Qy 337 YDSKTKNNDIALLMKLQKPLTENDLVKPVCLNPGMMLQPOLCWISGAGTEK-GKTSE 395
Db 271 Y---PKNDIALMKLQPLTFSGTVRPICLFPDPBELTPATPLWICGWFTKQNGKMSD 327
Qy 396 VLNAKAVLLIETORCNSRYVVDNLIITPAMICAGFLOGNVDSGCGSGPLVTSKNNIWL 455
Db 328 ILLQASQVQIDSTRCNADDAAYGGEVTEKMKCAGIPEGGVDTTCQDGGGLMYOSDQ-WHV 386
Qy 456 IGDTSGWGCACAYRPGVYGNVWFTDVIYQMRAD 491
Db 387 VGIVSWGYGCGSPSTFGVYTKVSAYLNNIYNWKAEE 422

RESULT 11

US-09-851-588-6
; Sequence 6, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; FILE REFERENCE: A-69108/DJ8/JUD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-851-588-6

FILE REFERENCE: A-68829-1/DJB/JJD/AMS
 CURRENT APPLICATION NUMBER: US/09/851,588
 CURRENT FILING DATE: 2001-09-24
 PRIOR APPLICATION NUMBER: US 09/642,252
 PRIOR FILING DATE: 2000-08-17
 PRIOR APPLICATION NUMBER: US 09/656,002
 PRIOR FILING DATE: 2000-09-06
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: Patent in version 3.1
 SEQ ID NO 6
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-851-588-6

Query Match 24.9%; Score 676.5; DB 4; Length 406;
 Best Local Similarity 39.1%; Pred. No. 1.9e-54;
 Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;
 QY 133 CDGVSHCPGGDENRCVRLY--GP-----NFILQVYSSQKSHHPVCCDDWNYG 181
 DB 46 CGELDCPLGDEEHCKVSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNTEALA 105
 QY 182 RAACRDNGY--KNFYSSQ-----GIVDDSGSTFVKLNTSAGNVDIYKLYHSDACS 232
 DB 106 ETACRQMGYSKPTFRAVEIGDPDQDLVDVVEITENSQELMRNMSG-----PCL 153
 QY 233 SKAVVSLRCIACGVNLNSRQSRIVGGESALPGAPWQVSLHVQNVHVCSSIIITPEWIV 292
 DB 154 SGLSVSLHCLACKSL---KTRPVVGGEEASVDSWPMQVSIQYDKQHVCGSILDPHWL 210
 QY 293 TAAHCVKELNPNHWTAFAGILR-QSPMFYAGAYQVEKVI---SHPNYDSKTKNDIAL 348
 DB 211 TAAHCFRKH-TDVFNVKVRAGSKLGSF---PSLAVAKIIIEFNPMY---PKNDIAL 262
 QY 349 MKLQKPLTFNDLVKPVCLPNPMMLOPEQLCWISGNGATEEK-GKTSEVLNAAKVLIIET 407
 DB 263 MKLQPLTFSGTVRPICLPFFDEELTPATPLWIGWFTKQNGKMSDILLQASVQVIDS 322
 QY 408 QRCNSRYVDNLITPAMICAGFLOGNVDSQCGSGPLVTSKNINWMLIGDTSWGSCAK 467
 DB 323 TRCNADDAIYQGEVTEKMMKAGIPGGVDTCCGDSGGPLMYQSDQ-WHVVGIVSWGCGG 381
 QY 468 AYRPGVYGNVMTDIYRQMRAD 491
 DB 382 PSTPGVYTKVSAYLNWYINVKAE 405

RESULT 12
 US-09-008-271A-6
 ; Sequence 6, Application US/09008271A
 ; Patent No. 6203979
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; Hillman, Jennifer L.
 ; Yue, Henry
 ; Guegler, Karl J.
 ; Corley, Neil C.
 ; Tang, Tom Y.
 ; Shah, Purvi
 ; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Dr.
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/008,271A
 FILING DATE: 16-Jan-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: <Unknown>
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Mohan-Peterson, Sheila
 REGISTRATION NUMBER: 41,201
 REFERENCE/DOCKET NUMBER: PF-0458 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-855-0555
 TELEFAX: 650-845-4166
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 435 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: COLNNOT13
 CLONE: 1337018
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 US-09-008-271A-6

Query Match 24.9%; Score 676.5; DB 3; Length 435;
 Best Local Similarity 39.1%; Pred. No. 2.1e-54;
 Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;
 QY 133 CDGVSHCPGGDENRCVRLY--GP-----NFILQVYSSQKSHHPVCCDDWNYG 181
 DB 75 CGELDCPLGDEEHCKVSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNTEALA 134
 QY 182 RAACRDNGY--KNFYSSQ-----GIVDDSGSTFVKLNTSAGNVDIYKLYHSDACS 232
 DB 135 ETACRQMGYSKPTFRAVEIGDPDQDLVDVVEITENSQELMRNMSG-----PCL 182
 QY 233 SKAVVSLRCIACGVNLNSRQSRIVGGESALPGAPWQVSLHVQNVHVCSSIIITPEWIV 292
 DB 183 SGLSVSLHCLACKSL---KTRPVVGGEEASVDSWPMQVSIQYDKQHVCGSILDPHWL 239
 QY 293 TAAHCVKELNPNHWTAFAGILR-QSPMFYAGAYQVEKVI---SHPNYDSKTKNDIAL 348
 DB 240 TAAHCFRKH-TDVFNVKVRAGSKLGSF---PSLAVAKIIIEFNPMY---PKNDIAL 291
 QY 349 MKLQKPLTFNDLVKPVCLPNPMMLOPEQLCWISGNGATEEK-GKTSEVLNAAKVLIIET 407
 DB 292 MKLQPLTFSGTVRPICLPFFDEELTPATPLWIGWFTKQNGKMSDILLQASVQVIDS 351
 QY 408 QRCNSRYVDNLITPAMICAGFLOGNVDSQCGSGPLVTSKNINWMLIGDTSWGSCAK 467
 DB 352 TRCNADDAIYQGEVTEKMMKAGIPGGVDTCCGDSGGPLMYQSDQ-WHVVGIVSWGCGG 410
 QY 468 AYRPGVYGNVMTDIYRQMRAD 491
 DB 411 PSTPGVYTKVSAYLNWYINVKAE 434

RESULT 13
 US-09-851-588-8
 ; Sequence 8, Application US/09851588
 ; Patent No. 6682890
 ; GENERAL INFORMATION:
 ; APPLICANT: Mack, David
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Wilson, Keith E.
 ; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
 ; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
 ; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
 ; CURRENT APPLICATION NUMBER: US/09/851,588
 ; CURRENT FILING DATE: 2001-09-24
 ; PRIOR APPLICATION NUMBER: US 09/642,252

;; PRIOR FILING DATE: 2000-08-17
;; PRIOR APPLICATION NUMBER: US 09/656,002
;; PRIOR FILING DATE: 2000-09-06
;; NUMBER OF SEQ ID NOS: 9
;; SOFTWARE: Patent in version 3.1
;; SEQ ID NO 8
;; LENGTH: 437
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-851-588-8

Query Match 24.9%; Score 676.5; DB 4; Length 437;
Best Local Similarity 39.1%; Pred. No. 2.1e-54;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;
QY 133 CDGVSCHPCGDEENRCVRLY--GP-----NFIQVYSSQKSWHFVQCDDWNNY 181
DB 77 CDGELDCPLGEDEHCHVKFPGRAVAVLSKDRSTLQVLDSATCNWFSACFDNTEALA 136
QY 182 RAACRMGY--KNNFVSSQ-----GIVDSGTSFPMKLNSTAGNVDIYKLVHSACS 232
DB 137 ETACROMGYSSKPTFRAVIGIPDQDLVVVEITENSQELMRNSSG-----PCL 184
QY 233 SKAVSLRGIACQVNLNSRQRIYVGSALPGAWPQVSLHVONVHVCGGSIITPEWIV 292
DB 185 SGLSLVHLCLACGKSL--KTRVVGGEASVDSWPNQVSIQYDKQVCGSILDPHWL 241
QY 293 TAACHVEKELNPNWHTAFAGILR-QSPFYAGYQVEKVI---SHPNYDSKTKNDIAL 348
DB 242 TAACHCRPRK--TDVFNKVKZAGSDKLSF---PSLAVAKIIIEFNPMY----PKNDIAL 293
QY 349 MKLQKLTNDLVKPCVCLNPGMLOPEQLCWISQWATEEK-KGTSVLNAAKVLLET 407
DB 294 MKLQFLTSGVTRPCLPFDEELTPATPLWICNGFTKQGGKMSDILLQASVQVIDS 353
QY 408 QRCNSRYVDNLITPAMICAGFLQNVDSQCGSGGLVTSKNNIWMIGTSGWGCAG 467
DB 354 TRCNADDAVCGEVTETKMCAGIPEGVDTCQDGGPLMYSDQ-WHYVGVISWGYCGG 412
QY 468 AVRPVGVYGVNMFVDIYRQMDAD 491
DB 413 PSTPGVYTKVSAYLNWYNNWAE 436

RESULT 14
US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 565566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinerdt, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170 X8574
TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-200-900A-2

Query Match 24.3%; Score 660; DB 1; Length 798;
Best Local Similarity 33.9%; Pred. No. 1.7e-52;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
QY 65 VCTQPSKSGTVCSKTKKALCIITL-----GTELVGAALAAAGLLWFKMGSKCSNSG 117
DB 374 VYTGPGVNDVFSITNMTVLFITDMIAKQGFKANFTTGVGLG-----IPEPCKEDN 426
QY 118 IECDSSTGCIINPSKWCQGVSHCPGEGEDENRCVRYG-----PNFIQVYSSQKSWH 169
DB 427 FQC-KDGEICPLVNLCDGFPCHCKDGSDEAHCVRLFNCTTDSGLVQFRIQ-----SIWH 479
QY 170 PVCQDDWNNYGRAACRDMGYKNNFYSSQGIIVDDSGTSPFKMLNTSAGNVDIYKLVHSD 229
DB 480 VACAENWTOISDDVCQLGLGTG--NSSVPTFTSGGPPYVNLNTAFNGSLI---LTPSQ 534
QY 230 ACSSKAVVSLRC--IACQVNLNSRQS-RIVGSSALPGAWPQVSLHVONVHVCGSII 286
DB 535 QCLEDSLILQCNKSCGKLVTOEVSFKIVGSDSREGAWPWVVALYFDDQVCGASLV 594
QY 287 TEWTVTAACHVCKEPNPNWHTAFAGILRQSFYAGYQVE---KVISHPNYDSKT 341
DB 595 SEDMLVSAACHVYGENMEPSKKAVALGLHMASNL---TSPQIETRLDQIVINPHYNKR 651
QY 342 KNDIALMKLOKPLTFNDLVKPCVCLNPGMLOPEQLCWISQWATEEKTEKTSVLNAAK 401
DB 652 KNDIAMHLEMKVNYDIQICLPEENOVFPFGRICSIAGMGALTYQGSADVLQEAD 711
QY 402 VLLIETQRCNSRYVDNLITPAMICAGFLQNVDSQCGSGGLVTSKNNIWMIGTSGW 461
DB 712 VPLISNEKQQQOMPEYN-IENWVCAGYEAGGVDSQCGDGGPLMCCQENNRWLLAGVT 770
QY 462 SSGCAKAYRPGVYGVNMFVDI 484
DB 771 GYQCALPNRPGVYARVFRFTWI 793

RESULT 15
PCT-US94-00616-2
; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-00616-2

Query Match		24.3%;	Score 660;	DB 5;	Length 798;
Best Local Similarity		33.9%;	Pred. No. 1.7e-52;		
Matches 150;		Conservative 67;	Mismatches 180;	Indels 46;	Gaps 12;
Qy	65	VCTQPKSPSGTCTSTKTKALCITLTL-----GTFLVGAALAGLLWKFMGSKCSNSG	117		
Db	374	VYTGPGFVNDVFTTNRWTVLFTDNMLAKQGFKANFTTGYGLG-----IPEPCKEDN	426		
Qy	118	IBCDSSGTCINPNWCDGVSHCPGGEDNRCVRLYG-----PNFILOVYSSQKSKWH	169		
Db	427	FQC-KDGECLPLVNLCDGFFHCKDGSDEAHCVLFFNGTTDSSGLVQFRIQ-----SIWH	479		
Qy	170	PVQDDWNNENYGRAACRDMGYKNFYSSQGI VDDSGSTSPFKMLNTSAGNVDIYKKLYHSD	229		
Db	480	VACAEWTTQISDDVCOLLGLGTG--NSSVPTSTGGPYNLTAPNGSLI---LTPSQ	534		
Qy	230	ACSSKAVVSLRC--IAGVNLNNSROS-RIVGESALPGAWPQVNSLHVQNVHVCGSII	286		
Db	535	QCLEDSLILLQCNYSKCGKXLTQEVSPKIVGGSDSREGAWPVVVALYFDDQVCGASLV	594		
Qy	287	TPEWITAAKCVKEPLNPNHWTAFAGILRQSPFMYGAGYQVE-----KVISHPNYDSKT	341		
Db	595	SRDLVSAAHCVYGRNMEPSKXAVLGLHMASNL---TSPQIETRLIDQIVINPHYNKER	651		
Qy	342	KNNDIALMKLQKPLTFNDLVKPVCLPNPMMQFPELQWISGWGATEEKGTSEVLNAAK	401		
Db	652	KNNDIAMVHLEMKVNYTDYIQICLPEENQVFPFGRICSIAGWGALIYQGSTADVLQEAD	711		
Qy	402	VLLIETORCNSRYVDNLIITPAMLCAGFLOGNVDSQGSQGGPLVTSKNNIWWLIGDTSW	461		
Db	712	VPLLSNEKCOQOMPEYN-ITENMVCAGYEAGGVDSQGSQGGPLMCOENRWLLAGVTSP	770		
Qy	462	GSQCAKAYRPGVYGNVMVFTDWI	484		
Db	771	GYQCALPNRPGVYARVPRFTEWI	793		

Search completed: June 1, 2004, 14:39:03
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:15:33 ; Search time 134 Seconds
(without alignments)
1158.470 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 2717
Sequence: 1 MALNSGSPPAIGPYENHGY.....VGNVMVFTDIYRQMRADG 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- SPTREMBL 25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2711	99.8	492	Q96T73	Q96T73 homo sapien
2	2121	78.1	490	Q7TN04	Q7TN04 mus musculus
3	2083	76.7	490	Q920K3	Q920K3 rattus norv
4	1009	37.1	767	Q9DGR2	Q9DGR2 xenopus lae
5	881	32.4	453	Q812A6	Q812A6 mus musculus
6	751	27.6	537	Q9BYE1	Q9BYE1 homo sapien
7	736.5	27.1	581	Q9BYE2	Q9BYE2 mus musculus
8	731	26.9	471	Q8CFE0	Q8CFE0 homo sapien
9	725.5	26.7	558	Q86YM4	Q86YM4 homo sapien
10	676.5	24.9	405	Q96886	Q96886 homo sapien
11	675	24.8	445	Q8CU17	Q8CU17 rattus norv
12	664.5	24.5	455	Q8CDR0	Q8CDR0 mus musculus
13	662.5	24.4	371	Q8CU16	Q8CU16 rattus norv
14	649	23.9	326	Q7Z280	Q7Z280 brachydanio
15	646.5	23.8	777	Q8CAN9	Q8CAN9 mus musculus
16	600.5	22.1	1111	Q80YN4	Q80YN4 rattus norv

17	567.5	20.9	624	6	Q95ME7	Q95ME7 oryctolagus
18	564.5	20.8	855	11	Q9DU17	Q9DU17 rattus norv
19	563.5	20.7	643	6	Q97506	Q97506 sus scrofa
20	560.5	20.6	624	11	Q9DAT3	Q9DAT3 mus musculus
21	558.5	20.6	422	4	Q8WVC1	Q8WVC1 homo sapien
22	553	20.4	624	11	Q91Y47	Q91Y47 mus musculus
23	544	20.0	1379	5	Q9V4N6	Q9V4N6 drosophila
24	543	20.0	845	13	Q9DGR1	Q9DGR1 xenopus lae
25	542	19.9	638	11	Q8ROP5	Q8ROP5 mus musculus
26	541	19.9	572	11	Q8BIK6	Q8BIK6 mus musculus
27	529	19.5	310	11	Q9QY29	Q9QY29 mus musculus
28	523	19.2	310	11	Q91XC4	Q91XC4 mus musculus
29	522.5	19.2	812	11	Q9ROW3	Q9ROW3 rattus norv
30	519.5	19.1	439	11	Q8BHM9	Q8BHM9 mus musculus
31	516	19.0	417	11	Q8VHJ4	Q8VHJ4 rattus norv
32	514	18.9	417	11	Q8VDV1	Q8VDV1 mus musculus
33	514	18.9	417	11	Q8VHK8	Q8VHK8 mus musculus
34	509.5	18.8	329	6	Q9GL10	Q9GL10 ovis aries
35	509.5	18.8	331	11	Q8RIA6	Q8RIA6 mus musculus
36	509.5	18.8	389	13	Q9PVX7	Q9PVX7 xenopus lae
37	508.5	18.7	279	11	Q7TNX3	Q7TNX3 mus musculus
38	508.5	18.7	417	11	Q8B210	Q8B210 mus musculus
39	506.5	18.6	331	11	Q8XU17	Q8XU17 mus musculus
40	505	18.6	328	11	Q8BUR6	Q8BUR6 mus musculus
41	504.5	18.6	257	11	Q8SZ04	Q8SZ04 mus musculus
42	504.5	18.6	320	13	Q7TDX2	Q7TDX2 xenopus lae
43	504.5	18.6	336	11	Q8OYD8	Q8OYD8 mus musculus
44	503.5	18.5	279	11	Q9QZ74	Q9QZ74 rattus norv
45	503.5	18.5	416	4	Q86T26	Q86T26 homo sapien

ALIGNMENTS

RESULT 1

Q96T73	PRELIMINARY; PRT; 492 AA.
ID	Q96T73
AC	Q96T73;
DT	01-DEC-2001 (TREMBlrel. 19, Created)
DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT	01-CCT-2003 (TREMBlrel. 25, Last annotation update)
DE	Epitheliasin.
GN	TPRS82.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21223025; PubMed=11322890;
RA	Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,
RA	Hoidal J.R.;
RT	"Cloning and characterization of the cDNA and gene for human
RT	epitheliasin."
RL	Eur. J. Biochem. 268:2687-2699 (2001).
CC	-L- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR	EMBL; AF329454; AAKS3559.1; -
DR	HSSP; P00761; IANL.
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.
DR	GO; GO:0008233; F:peptidase activity; IEA.
DR	GO; GO:0005044; F:scavenger receptor activity; IEA.
DR	GO; GO:0004295; F:trypsin activity; IEA.
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR	InterPro; IPR009003; Cys_Ser_trypsin.
DR	InterPro; IPR002172; LBL_receptor_A.
DR	InterPro; IPR01254; Peptidase S1.
DR	InterPro; IPR01314; Peptidase S1A.
DR	InterPro; IPR01190; Strc_receptor.
DR	Pfam; PF00089; trypsin_1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	SMART; SM00192; LDLA; 1.
DR	SMART; SM00202; SR; 1.

DR	SMART; SMO0020; Tryp_Spc; 1.
DR	PROSITE; PS01209; LDLRA_1; 1.
DR	PROSITE; PS0068; LDLRA_2; 1.
DR	PROSITE; PS0287; SRCR_2; 1.
DR	PROSITE; PS0240; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
KW	Hydrolase; Protease; Serine protease
SQ	SEQUENCE 492 AA; 53863 MW; 3ABA755BF276DADF CRC64;

Query Match	99.8%; Score 2711; DB 4; Length 492;
Best Local Similarity	99.6%; Pred. NO. 1.1e-232;
Matches 490; Conservative	2; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MALNGSSPPAIGPYENHGYPENPYPAQPTVTVTFVEVHPAOYYSPPVPAPRVLTQA	60
Dd	1	MALNGSSPPAIGPYENHGYPENPYPAQPTVTVTFVEVHPAOYYSPPVPAPRVLTQA	60
Qy	61	SNPVTCTQKSPSGTVCSTKKALCITLTLGTFLVGAAALAGLLWKFMGSKCSNSGIIC	120
Dd	61	SNPVTCTQKSPSGTVCSTKKALCITLTLGTFLVGAAALAGLLWKFMGSKCSNSGIIC	120
Qy	121	DSSGTCINSNWCDGVSHCPGEDENRCVRILYGPNFLOLVYVSQRKSWHPVCODDNWNY	180
Dd	121	DSSGTCINSNWCDGVSHCPGEDENRCVRILYGPNFLOLVYVSQRKSWHPVCODDNWNY	180
Qy	181	GRAACKDMGYKNFYSSQGIIVDDSGTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR	240
Dd	181	GRAACKDMGYKNFYSSQGIIVDDSGTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR	240
Qy	241	CACGVNLNNSQSRIYGGESALPGAWPWQVSLHVQNHHVCGGSIIIPWIVTAACHCEVK	300
Dd	241	CACGVNLNNSQSRIYGGESALPGAWPWQVSLHVQNHHVCGGSIIIPWIVTAACHCEVK	300
Qy	301	PLNNPHWTAFAGILLRQSPMFYAGYQVEKIYHSNPYDSKTNNNDIALMKLKQLPTFNDL	360
Dd	301	PLNNPHWTAFAGILLRQSPMFYAGYQVEKIYHSNPYDSKTNNNDIALMKLKQLPTFNDL	360
Qy	361	VKEPVCLPNFGMLLPOLCWISGWATSEKGTSEVLNAAKVLIIETQRCNSRYVDNLI	420
Dd	361	VKEPVCLPNFGMLLPOLCWISGWATSEKGTSEVLNAAKVLIIETQRCNSRYVDNLI	420
Qy	421	TPAMI CAGFLQGNDVSCQDGGPLVTISKNNIWWLIGDTSMGSCCAKAYRPYGVNMVF	480
Dd	421	TPAMI CAGFLQGNDVSCQDGGPLVTISKNNIWWLIGDTSMGSCCAKAYRPYGVNMVF	480
Qy	481	TDWIYQMRADG 492	
Dd	481	TDWIYQMRADG 492	

```

RESULT 2
Q7TN04          Q7TN04          PRELIMINARY;          PRT;          490 AA.
AC      Q7TN04;
AD      01-OCT-2003 (TtEMBLrel. 25, Created)
DT      01-OCT-2003 (TtEMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TtEMBLrel. 25, Last annotation update)
DE      Tmpres2 protein.
DS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_taxid=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=129; TISSUE=Breast tumor;
RX      MEDLINE=22386257; PubMed=12477932;
RA      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Alschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

```

Query Match	78.1%	Score 2121	DB 11	Length 430
Best Local Similarity	78.4%	Pred. No. 3.7e-180		
Matches 385	Conservative 43	Mismatches 61	Indels 2	Gaps 2
Qy/	1	MALNSGSPPAIGPYENHGYOPENPYPAQQTWPTVYVEHPAQPYPSPVQYAPRVLTAQ	60	
Db	1	MALNSGSPPGICPYENHGYQSEHI CPPRPVAPNGYNLYPAQYPSPVQYAPRITTAQ	60	
Qy	61	SNPVCTQPKSPGTVCTSKTKALCITLITGLTFPLVGAALAGLLWFKMGSKCSNSGIEC	120	
Db	61	STSVITHTPKS - SGALCTSKSKSLCALAGTGLTGAATAVALLWRFWDNSCSTSENEC	119	
Qy	121	DSSGTCINPNWCDGVSHCPGSGDENRCVLYGNFILQVYSSQKSWHPVCQDDWNNY	180	
Db	120	GSSGTCISSLWCDGVAHCPGSDENRCVLYGSGFILQVYSSQKAWYPVCQDDWSEY	179	
Qy	181	GRAACRMGYKNPFPYSSGGIVDDSGTSPFKLNTSAGNVLTKYLYHSDACSSKAVVSLR	240	
Db	180	GRAACKDMGYKNPFPYSSGGIPDQSGATSPFKLNVSSGNVDLYKLYHSDCSSRNWVSLR	239	
Qy	241	CIACGVNLNRSRQRIYGGESALPGAWPQVSLHVQNVHVGCGSGIITPEWITVTAACVCK	300	
Db	240	CIECGVR - SVKQSRIVGGLNASFGDPWQVSLHVQGVHVGCGSGIITPEWITVTAACVCE	298	
Qy	301	PLNNPMTWTFAGILRQSPFYGAGYQVEKVISHPNVDTSKTKNDIALMKLQKPLTNDL	360	
Db	299	PLSSFRYWTAFAGILRQSLPFPYGRHQVEKVISHPNVDTSKTKNDIALMKLQPLAFNDL	358	
Qy	361	VKPVCLNPGMMLPEOLCWISGNGATTEKGTSEVLNAKVLLETQRCNSRYVDNLI	420	
Db	359	VKPVCLNPGMMLDDEQWISGNGATTEKGTSEVLNAKVLLETQRCNSRYVDNLI	418	
Qy	421	TPAMICAGFLQGNVDSQCGSGGGLVTSKNIIWLLIGDTSWGSCKAAYRGVYGNVMVF	480	
Db	419	TPAMICAGFLQGSVDQCGSGGGLVTLKNGIWWLLIGDTSWGSCKAALREPGVYGNVTVF	478	
Qy	481	TDWIYQWVRAD 491		
Db	479	TDWIYQWVRAN 489		
RESULT 3				
Q920K3				
ID	Q920K3	PRELIMINARY;	PRT;	490 AA.
AC	Q920K3;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	TMPSR2.			
OS	Rattus norvegicus (Rat).			

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Tsuzuki S.;
RT "TPRS2, Rat.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -/- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB073550; BAB70683.1; -.
DR HSP; P00761; IANL.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001272; LDL_receptor A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; S1cr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00200; Tryp_SPC; 1.
DR PROSITE; PS00068; LDLA_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease.
SQ SEQUENCE 490 AA; 53518 MW; 2BC691551CAC409A CRC64;

Query Match 76.7%; Score 2083; DB 11; Length 490;
Best Local Similarity 77.0%; Pred. No. 9e-177;
Matches 378; Conservative 47; Mismatches 64; Indels 2; Gaps 2;

Qy 1 MALNSGSPFAIGPYENHGYQENPYPAQTPVTVYVHPAQYPPFPQYAPRVLTQA 60
Db 1 MALNSGSPFGIPYENHGYSEHVSPRPVPSGYNLYPAQSCPSFPQYAPRVLTQA 60
Qy 61 SNFVCTOPKSPSTGVTCTSKKALCTLTGLTVGAALAGLLWKPMGSKNSGIEC 120
Db 61 STPAIHQIPRS-SGLTCTSKSKMLVALAGLSPVAANAAGLLWKFWDKSCSSSEMEC 119
Qy 121 DSSGTCINPNSWCDGVSHCPGEGDENRCVRLYGPNFILQVYSSQKSWHPVQCDDMNENY 180
Db 120 GSSGTCISSLWCDGVAQCPNGKDNRCVRLYGTSTFTLQVYSSQKAWYPVQCDDMNESY 179
Qy 181 GRAACRDGMGNKXNFYSSQGIYVDDSGSTSPFMKLNAGNVDIYKLYHSDACSSKAVVSLR 240
Db 180 GRAACKDMGYNKXNFYSSQGIYVDDSGSTSPFMKLNAGNVDIYKLYHSDACSSRMVYSLR 239
Qy 241 CIACGVNLNSRQSRIVGSGALPGAWPQVSLHVQNVHVCVGGSIITPEWITVAHCVCK 300
Db 240 CIECQVR-SVARQSRIVGSGALPGAWPQVSLHVQNVHVCVGGSIITPEWITVAHCVCK 298
Qy 301 PLNNPWHMTAFAGILRQSFYMGYQVEKVISHPNYDSKTKNDIALMKLQKPLTNDL 360
Db 299 PLSSPRYWTAFAGILKQSLMFYGRHQVEKVISHPNYDSKTKNDIALMKLQKPLTNDL 358
Qy 361 VKPVCPLNPGMLOPEQLCWISGWGATEEKGKSEVLNAAKVLIIETQCNRSRVYDNLJ 420
Db 359 VKPVCPLNPGMLOPEQLCWISGWGATEEKGKSEVLNAAKVLIIETQCNRSRVYDNLJ 418
Qy 421 TPAMTCAQFLOGNVDSQCGDGGPLVTSKNNIWMILGDTSGWGCACAYRPGVGNVYVF 480
Db 419 TPAMTCAQFLOGNVDSQCGDGGPLVTLKNEIWMILGDTSGWGCACAYRPGVGNVYVF 478
Qy 481 TDWIIYQMRAD 491
|||||:|||||:

Db 479 TDWIIYQMRAN 489
RESULT 4
Q9DGR2 PRELIMINARY; PRT; 767 AA.
ID AC Q9DGR2;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Embryonic serine protease-2.
GN XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20363741; PubMed-10903452;
RT Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
from Xenopus laevis.";
RL Gene 252:209-216(2000).
CC -/- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB038497; BAB08217.1; -.
DR HSP; P00766; 1CHG.
DR MEROPS; S01.049; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 8.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLA_1; 8.
DR PROSITE; PS0068; LDLA_2; 2.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 96001 MW; E0566A38796DE96E CRC64;

Query Match 37.1%; Score 1009; DB 13; Length 767;
Best Local Similarity 50.9%; Pred. No. 7.2e-81;
Matches 189; Conservative 49; Mismatches 131; Indels 2; Gaps 2;

Qy 120 CDSGTCINPNSWCDGVSHCPGEGDENRCVRLYGPNFILQVYSSQKSWHPVQCDDMNEN 179
Db 395 CGSSVSVLLSQWCDGVSDCPYGEDEMSCVSLYPADFLQVYSTSVSAWLPVCSYWNDD 454
Qy 180 YGPAACRDGMGNKXNFYSS-QGIVDDSGSTSPFMKLNAGNVDIYKLYHSDACSSKAVV 238
Db 455 FGFAQCDGFGNGSSVNRVYDTLMSYPAPNGYFKLISGYWRSPFYTSVQYSSYSGNV 514
Qy 239 LRCIACGVNLNSRQSRIVGSGALPGAWPQVSLHVQNVHVCVGGSIITPEWITVAHCV 298
Db 515 LHCISCGVS-NNSLSVSRIVGGTFAVLGNWQVNLQYITGVLCGGSIISPKMIVTAAHCV 573
Qy 299 EKPLNPFWTAFAGILRQSFYMGYQVEKVISHPNYDSKTKNDIALMKLQKPLTNDL 358
Db 574 YGSYSASGRVFPAGTLTTPSYNNAAYFVEKVIHHPGYKSTYNDIALMKLQKPLTNDL 633
Qy 359 DLVKPVCPLNPGMLOPEQLCWISGWGATEEKGKSEVLNAAKVLIIETQCNRSRVYDNL 418
Db 634 YTTQVCLPNSGMFWAGTTTISGWSYEGSVSTYLYAAIPLIDSNVCNQSYVANG 693
|||||:|||||:


```
QY 419 LITPAMICAGFLQGNVDSQCGDGGPLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNM 478
Db 694 QITSSMICAGYLSGGVDTCQDGGPLVYKRNKGTWHLVGDTSWGGDCARANKFVGYNVT 753
QY 479 VFTDWIYRQMR 489
Db 754 TFLWYISQMR 764

RESULT 5
Q812A6 PRELIMINARY; PRT; 453 AA.
ID Q812A6
AC Q812A6;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Transmembrane proteinase tmp:ses3.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rao N.V., Rao G.N., Hoidal J.R.;
RT "Genomic Organization of Murine Transmembrane Proteinases.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF479687; AA033581.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS0068; LDLRA_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 453 AA; 49505 MW; 1BE7ECD6CB3DD894 CRC64;

Query Match 32.4%; Score 881; DB 11; Length 453;
Best Local Similarity 43.1%; Pred. No. 8.8e-70;
Matches 188; Conservative 63; Mismatches 139; Indels 26; Gaps 12;

QY 69 PKSPSGTCTSKTKKALCI-----TLTLGTFVLVGAALAGLWKPWGSKNSGIECDSSG 124
Db 28 PVAPDGAQAQILSLPLKFFPIIVIGIILALALATGLIHF---DCSGK-YRCHSSF 83

QY 125 TCINPWNCDGVSHCPGSGEDENRCVRLGPNFVLQVTSQKSWHPVCQDDWNNYGRAA 184
Db 84 KCIELTARCDGVSDCKNAEDYRCVRVSGQRAALQVFTA--AAWTWCSDDKSHYAKIA 141

QY 185 CRDMGYKNNFYSSQGI-----VDDSGSTSMKLNTSAGNVDIYKLYHS----DACSKAV 236
Db 142 CAQLGFP-SYVSSDLRLVDALDEEQFGQFVSIHLLSD-DKVTALHSHVYNNREGCTSGHV 199

QY 237 VSLRCIAGVNLNSRQRVIGGESALPGANPWQSLHVQNVHVCQGSIIIPMIVTAAH 296
Db 200 VTLKCSACGTRIGYS--PRIVGGNNSSLTQMPWQVSLQFQGYHLCCGSIIPLMIVTAAH 257

QY 297 CVEKPLNNPWHWTFAPILRQSFMPYGA-GYQVEKVI SHPNYDSKTKNNNDIALMKLQKPL 355
```

```
Db 298 CV-YDLXHPKSWTVQVGLV--SLNDSVPVSHLVEKIIYHKKYKPKRLGNDIALMKLSEPL 314
QY 356 TFDNLVFPVCLPNPQMMLQPEQLCWIISGWGATEEKGTSEVLNAAKVLIIETORCNRYY 415
Db 315 TFDETIQICLPNSENFPDGKLCWTSWGATEDGGDASPVLNHAAVPLISNKICNHRDV 374
QY 416 YDNLITPAMICAGFLQGNVDSQCGDGGPLVTSKNNIWWLIGDTSWGGCAKAYRPGVYG 475
Db 375 YGGIISPSMLCAGYLKGGVDSQCGDGGPLVTCQRRRLWKLVGATSGFGICCAEVNKPQVYT 434
QY 476 NMVFTDWIYRQMRAD 491
Db 435 RITSFLDWIHEQLERD 450

RESULT 6
Q9BYE1 PRELIMINARY; PRT; 537 AA.
ID Q9BYE1
AC Q9BYE1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Mosaic serine protease.
GN MSPS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmin S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
RL without a transmembrane domain from human lung.";
RL Biochim. Biophys. Acta 1518:204-209 (2001).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB048797; BAB39742.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.087; -.
DR GO; GO:0016021; C:integral to membrane; NAS.
DR GO; GO:000508; P:proteolysis and peptidolysis; NAS.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 537 AA; 58102 MW; A39FF4E8816DAECF CRC64;

Query Match 27.6%; Score 751; DB 4; Length 537;
Best Local Similarity 32.0%; Pred. No. 4.2e-58;
Matches 173; Conservative 81; Mismatches 198; Indels 88; Gaps 14;

QY 5 SSSPPAIGYYENHGVQENPYPAQTVVPT-VYEVHPAQYVPSVP----- 50
Db 27 ACTPPORA-----SFAQASPAQSPAGTTPGRASPAQSPAGTTPGRASPGRASPAQ 78

QY 51 ----QYAPRLVTOA-SNPVVCTQPKSPSQ-----TVCTSKTKKALCITITLGTFLVG 97
Db 79 ASPAQASPAQASPARASPALASLSRSSGSSSSASASVTTSTPTRVLYVREATPVGAVPTR 138
```

QY 98 AALA-----AGLLWKFMSKCSNSGIECDSSGTCINPSNMCQGVY 137
 Db 139 SSPARSAPATRAATRESVPQFQCHGTGIRYKEQRESCPKHVR-----CDGVY 185
 QY 138 HCPGGEENRCVLYGKFNFIQVYSSORKSWHPVQCDNNENYGRAACRDMGYKXNFYSS 197
 Db 186 DCKLSDELGVRFWDKSLKTIYSGSSHQWLPICSSNNWDSYSEKTCRQLQFESAHRRT 245
 QY 198 Q-GTVDSGSGTSFMKLNSTAGNVDIYKLYHSDACSKAVVSLRCLACGNLNSRSQRI 256
 Db 246 EVAHRDPANFESILRYNST-----IQESLHRSCHPCSORVISLQCSHGCLR---AMTGR 296
 QY 257 VGGESALPGAMPWQVSLHVQNVHVCSSGIITPFWITAAHCV-----EKPLNPNHWTAFAGILRQSEM 312
 Db 297 VGGALADSQWPCVSLHFGTHICGGTLIDAQWLVTAACHCFVTRKVLG---WKVYA 353
 QY 313 GILRQSFMYCAGYQVKEVISHPNYDSKTKNDIALMKLQKPLTENDLVKPCLEPNQMW 372
 Db 354 GTSNLHQLPEAA--SIAEIIINSNYTDEDDYDIALMRSLKPLTSAHHPACLPHGQT 411
 QY 373 LQPEQLCNIWSGATEE-KGKTSEVNAKVLITETORNSRYVYDNLITPAMICAGFLQ 431
 Db 412 FSLNETCWTGFGKTRTDDKTSPLREVQVNLIDFKKNDYLVDSYLTPTMWCAGDLH 471
 QY 432 QNVDSGCGSGPLVTSKNNIWLIGDTSWGSCKAKAYRFGVYGVNVTFTDIYQMRAD 491
 Db 472 GGRDSCQDSGGPLVCEQNNRWLAGVTSWGTGCGQRNKGVTYKTVTEVLPWIYSKM 531

RESULT 7
 Q9BYE2 PRELIMINARY; PRT; 581 AA.
 ID Q9BYE2
 AC Q9BYE2
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Membrane-type mosaic serine protease.
 GN MSPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=21167393; PubMed=11267681;
 RA Kim D.R., Sharmir S., Inoue M., Kido H.;
 RT "Cloning and expression of novel mosaic serine proteases with and
 RT without a transmembrane domain from human lung.";
 RL Biochim. Biophys. Acta 1518:204-209 (2001).
 DR EMBL; AB048796; BAB39741.1; -;
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.087; -;
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDua; 1.
 DR SMART; SM00202; Tryp_Spc; 1.
 DR PROSITE; PS00287; SRCR 2; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 581 AA; 62689 MW; 4DABE24D7D5BA44 CRC64;

Query Match 27.1%; Score 736.5; DB 4; Length 581;
 Best local similarity 22.3%; Pred. No. 9.1e-57;
 Matches 171; Conservative 78; Mismatches 198; Indels 82; Gaps 14;
 QY 5 SGSPATGPIYENHGYQENPYPAQPTVTVVYVHVAQYYP----- 46
 Db 62 AGTPPGRA-----SPGRASPAQAS-----PARASPALASLSRSSSSGRSSARS 104
 QY 47 -----SPVPOYAPRVLTAQSNPVVCTQPKSPGTCVTKTKKALCITLT-----L 91
 Db 105 ASVTTSPTVYLVRAVPGVPIRSSPARSAPATRAATRESPTSLPKFTWREGQKQLPLI 164
 QY 92 GTFLVGAALAAAGLL-----WKFMGSKCSNSGI-ECDSSGTCINPSNMCQGVHCPGGEDE 145
 Db 165 GCVLIIALVLSIIILFQFWQ-----GHTGIRHKEQRESCPKHVRCDGVVDCCLKSDE 218
 QY 146 NRCVRLVGPFIQVYSSQRKSWHPVQCDNNENYGRAACRDMGYKXNFYSSQ-GIVDDS 204
 Db 219 LGCVRFWDKSLKTIYSGSSHQWLPICSSNNWDSYSEKTCRQLQFESAHRITTEVAHRDA 278
 QY 205 GSTSFMKLNSTAGNVDIYKLYHSDACSKAVVSLRCLACGNLNSRSQRIYVGGESALP 264
 Db 279 NSFSILRYNST-----IQESLHRSCHPCSORVISLQCSHGCLR---AMTGRIVGGALAD 329
 QY 265 GAMPWQVSLHVQNVHVCSSGIITPFWITAAHCV-----EKPLNPNHWTAFAGILRQSEM 320
 Db 330 SKWPMQVSLHFGTHICGGTLIDAQWLVTAACHCFVTRKVLG---WKVYAGTSMHLQ 386
 QY 321 FYGAGYQVKEVISHPNYDSKTKNDIALMKLQKPLTENDLVKPCLEPNQMWLQPEQLCW 380
 Db 387 PEAA--SIAEIIINSNYTDEDDYDIALMRSLKPLTSAHHPACLPHGQTFSLNETCW 444
 QY 381 ISGWGATEE-KGKTSEVNAKVLITETORNSRYVYDNLITPAMICAGFLQVDSGCG 439
 Db 445 ITGFGKTRTDDKTSPLREVQVNLIDFKKNDYLVDSYLTPTMWCAGDLHGGDRSCG 504
 QY 440 DSGGGLVTSKNNIWLIGDTSWGSCKAKAYRFGVYGVNVTFTDIYQMR 488
 Db 505 DSGGGLVCEQNNRWLAGVTSWGTGCGQRNKGVTYKTVTEVLPWIYSKM 553
 RESULT 8
 Q8CFEO PRELIMINARY; PRT; 471 AA.
 ID Q8CFEO
 AC Q8CFEO
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to mosaic serine protease (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
 RA Strauberg R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC042878; AAB42878.1; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00530; SRCR; 1.
 DR Pfam; PF00089; trypsin; 1.

DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1
DR SMART; SM00202; TRYP_SPC; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
FT NON ITR
SQ SEQUENCE 471 AA; 52535 MW; ED58CFB6B7C3BCC4 CRC64;

Query Match 26.9%; Score 731; DB 11; Length 471;
Best Local Similarity 35.8%; Pred. No. 2.1e-56;
Matches 166; Conservative 75; Mismatches 183; Indels 44; Gaps 13;

QY 47 SPVPEQYAPRVLTQASNPVCTQKPSGTVCT-----SKTKKAL-----CITL 89
Db 19 SPTRVVLVRATPGVAPIRASPARSAPATRAFPSPGLSPFKFSQETQRLPLIGCVIL 78

QY 90 TLGTFLVGAALAAAGLLWKMGKCSNSGIEC-DSSGTCINPNSWCDGVSHCPGGEDNRC 148
Db 79 -----LISLVISILLIFYFWR---GHTGIKYEPLSECFIHAVRCGDGVDCMKSDLGC 130

QY 149 VRLYGNFILQVYSSQKSWHPVCDQDNNENYGRAACRDMGYKNNFYSSQGIIVDSGTS 208
Db 131 VRFDWKLSLLKYVSGSGEWLPVCSNNWDTDSKRTCOQLGF-DSAYETTEVAHRDITSS 189

QY 209 FKLNTSAGNVDIYKLYHSDACSKAVVSLRCLACGVNLNNSRQSRIVGGESALPAMP 269
Db 190 FL---LSEYNTTIQESLYRSQ-CPERRYVSLQSCHEGLR---AMTGRIVGGALLTSES 242

QY 269 MQVSLHVQNVHVCVGGSIITPEWIVTAACHV-----EKPLNPNWHTAPAGILRQSF 324
Db 243 MQVSLHFGTHICGGTILDAQWVLTAAHCFVTRKLEGG---WKYVAGTSLNHLQLEA 299

QY 325 GYQVEKIVSHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMLQPEQLCWISGW 384
Db 300 --SISQIIINGNTYDEQDDYDIALRLSKPLSLSAHHPACLPMEHGTQFGLNCTWTGF 357

QY 385 GATEERK-KTSEVLMAAKVLLIETORCNSRVYDNLITPAMICAGFLQGNVDSQCGSGG 443
Db 358 GKTKEDEKTSFPLREVQVNLIDFKKNDLVYDSYLTFRMWCAGDLRGHSDSCQSGSG 417

QY 444 PLVTSKNTWLLIGTWSGSGCAKAYRPGVGNVGVFTDVIYRQMRAD 491
Db 418 PLVCEQNRYLAGVTSWGTGCGQKRNKPGVYTKVTEVLPFIYRKMESE 465

RESULT 9
Q86YM4 PRELIMINARY; PRT; 558 AA.
AC Q86YM4; PRELIMINARY; PRT; 558 AA.
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Transmembrane protease serine 6.
GN TWPRSS6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP Park T.J.; Park W.J.;
RT "Homo sapiens transmembrane protease, serine 6 (TWPRSS6) mRNA."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV190317; AAC38062.1;
DR GO; GO:0016020; Cmembrane; IEA.
DR GO; GO:0004263; Fchymotrypsin activity; IEA.
DR GO; GO:0008233; Fpeptidase activity; IEA.
DR GO; GO:0005044; Fscavenger receptor activity; IEA.
DR GO; GO:0004295; Ftrypsin activity; IEA.
DR GO; GO:0006508; Pproteolysis and peptidolysis; IEA.

DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Scr_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; TRYP_SPC; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
SQ SEQUENCE 558 AA; 60432 MW; 4AC817FCD70D7017 CRC64;

Query Match 26.7%; Score 725.5; DB 4; Length 558;
Best Local Similarity 32.2%; Pred. No. 8.2e-56;
Matches 171; Conservative 81; Mismatches 197; Indels 82; Gaps 15;

QY 5 SGSPALGPGYENHGYQENPYPAQPTVVPVYVNHQAQYYP----- 46
Db 62 ACTPPGPA-----SPGRSPAQAS-----PARASPALASLSRSSSSGRSSARS 104

QY 47 -----SPVPQYAPRVLTQASNPVCTQKPSGTVCTSKTKKALCITLT-----L 91
Db 105 ASVTTSPTRVVLVRATPGVAPIRSSPARSAPATRAFPSPGLSPFKFWREGOKQLPLI 164

QY 92 GTFLVGAALAAAGLL-----WKFMGSKCSNSGIEC-DSSGTCINPNSWCDGVSHCPGGED 145
Db 165 GCVLIIILVLSLILFQFWC-----GYTGIRYKEQRESCEPEHARRDGVDCIKLSDE 218

QY 146 NCVSLYGFNFILQVYSSQKSWHPVCDQDNNENYGRAACRDMGYKNNFYSSQ-GIVDDS 204
Db 219 LCVRFDWKLKLLKYVSGSGHWLPICSSNNWDSYSEKTCQQLGFSAHRTTEVAHRDFA 278

QY 205 GSTSPKMLNTSAGNVDIYKLYHSDACSKAVVSLRCLACGVNLNNSRQSRIVGGESALP 264
Db 279 NFSFTLYRNTST-----IOESLHSE-CPSORVYSLQSCHEGLR---AMTGRIVGGALL 329

QY 265 GAWPQVSLHVQNVHVCVGGSIITPEWIVTAACHV-----EKPLNPNWHTAPAGILRQSF 320
Db 330 SWPQVSLHFGTHICGGTILDAQWVLTAAHCFVTRKLEGG---WKYVAGTSLNHLQ 386

QY 321 FYGAGYQVEKIVSHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMLQPEQLCW 380
Db 387 PEAA--SIAETIINSNTYDEDDYDIALMRLSKPLSLSAHHPACLPMEHGTQFSLNETCW 444

QY 381 ISGWGATEE-KGKTSEVLMAAKVLLIETORCNSRVYDNLITPAMICAGFLQGNVDSQCG 439
Db 445 ITGFKETRETDDKTSPFLREVQVNLIDFKKNDLVYDSYLTFRMWCAGDLRGHSDSCQ 504

QY 440 DSGGLPVTSKNNIWLIGTWSGSGCAKAYRPGVGNVGVFTDVIYRQMR 490
Db 505 DSGGLPVTCEQNRYLAGVTSWGTGCGQKRNKPGVYTKVTEVLPFIYRKMESE 555

RESULT 10
Q96E86 PRELIMINARY; PRT; 405 AA.
AC Q96E86; PRELIMINARY; PRT; 405 AA.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

RESULT 11	
Q8CJU17	
ID	PRELIMINARY; PRT; 445 AA.
AC	Q8CJU17;
DT	01-MAR-2003 (TREMBlurel. 23, Created)
DT	01-MAR-2003 (TREMBlurel. 23, Last sequence update)
DT	01-OCT-2003 (TREMBlurel. 25, Last annotation update)
DE	Adrenal mitochondrial protease long variant.
GN	AMP.
OS	Rattus norvegicus (Rat).
CC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

RESULT 12	
Q8CDRO	PRELIMINARY; PRT; 455 AA.
ID	
Q8CDRO	
AC	
Q8CDRO;	
DT 01-MAR-2003	(TREMELrel. 23, Created)
DT 01-MAR-2003	(TREMELrel. 23, Last sequence update)
DT 01-MAR-2003	(TREMELrel. 23, Last annotation update)
DT 01-OCT-2003	(TREMELrel. 25, Last annotation update)
DT 01-OCT-2003	(TREMELrel. 25, Last annotation update)

GN SI:D269610.3.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Corby N.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL672083; CAD61105.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00668; LDLRA_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Transmembrane; Protease.
FT NON TER 1
SQ SEQUENCE 326 AA; 35561 MW; 8D6F12214393CDB2 CRC64;

Query Match 23.9%; Score 649; DB 13; Length 326;
Best Local Similarity 36.9%; Pred. No. 2.6e-49;
Matches 140; Conservative 50; Mismatches 121; Indels 68; Gaps 8;

QY 110 GSKCSNGIECDSSGTCINPSNWCDCGVSHCPGEGEDNRCVRLYGPNFILQVYSSORKSWH 169
DB 1 GLRSCSGKFCVSVSRVCRISRNACVDCGVQDCRDELCNCRVSGSHVLOVFG--RGLWR 58

QY 170 PWCQDDWNYGPEAAACRDGKXNFYSSQGIYDDSGSTFPMKNTSAGNVDIYKLYHSD 229
DB 59 TVCSGWDGSLTLACRQLGY----- 79

QY 230 ACSKAVLSLRICIACGNLNSRSRQSRIVGSEALPGAMPQVSLHVQNVHVGCGSIITPE 289
DB 80 ---SRSAIS-----SRIVGNVSKSGVQVPSLHYQNYLQCGSIIS 121

QY 290 WIVTAACHVEKPLNPNHWTAPAGILRQSFMYGAGYQVEKVLISHPNYDSKTKNNDIALM 349
DB 122 WILTAACHV--FGPAQVPLMDVYAGLNLPL--LSKAEHRSVEKIIYHANFRSKSFSYDIALI 179

QY 350 KLOKPLTFNDLVKPLCPNPGMMLQPEQLCWISGWCATBEKGTSEVLNAAKVLLIETOR 409
DB 180 KLLPLTFNDLQIAPICLPNKGESFGKQMCCLISGWCATVDSGETSLSLHVAQVPLLSNKE 239

QY 410 CNSRYVNDLIITPAMICAGPLQNVDSQCQDGGPLVTSKNNIWMILGTSWGSCKAY 469
DB 240 CRKLG-----LTNNVCTEFLRG-VGTQCQDGGPL-ACQGSALTIVGTGSDWDCGKYN 292

QY 470 RGVYGNVYFTDIYRQM 488
DB 293 KPGIYTSISEALTWIOEQM 311

RESULT 15
Q8CAN9

ID Q8CAN9 PRELIMINARY; PRT; 777 AA.
AC Q8CAN9
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Protease (Fragment).
GN PRSS7 OR A130097D21RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=22354683; PubMed=12466851;
EX The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK038356; BAC29973.1; -.
DR MGD; MGI:1197523; Pss87.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00859; CUB.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000998; MAM_domain.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00137; MAM; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00668; LDLRA_2; 1.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS00660; MAM_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
FT NON TER 1
SQ SEQUENCE 777 AA; 87314 MW; A18E2P4ECF06D3A8 CRC64;

Query Match 23.8%; Score 546.5; DB 11; Length 777;
Best Local Similarity 30.5%; Pred. No. 1.4e-48;
Matches 164; Conservative 92; Mismatches 196; Indels 85; Gaps 20;

QY 18 HGYPENPYPAQPTWPT-----VVEV-HPAQYVPSVPQYAPRVLTQAS---- 61
DB 253 NGICSQSPY-EPILVTFPPPELFTDCGGPFELWEPNSTSS--PNFPDKYPNQASCIWN 309

QY 62 -----NPV-----VCTQPKSPSGIVCTSKTKAL 85
DB 310 LNAQRGKNIQLHFQFDFLENINDVVEVRDGGFDFSLLLAVYTG- GFPKDLFTSTNRMTV 368

QY 86 CITLTGTFIVG--AALAAGLLWKFMG--SKCSNGIECDSSGTCINPSNWCDCGVSHCPG 141
DB 369 IFTTNMTRRGKXANFTSGY---YLGIPBPCQDDEFQC-KDGNICPLGNLCDSVPHCRD 424

Search completed: June 1, 2004, 14:37:30
Job time : 136 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 1, 2004, 14:45:30 ; Search time 3778 Seconds
(without alignments)
3988.881 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 2717
Sequence: 1 MALNSGPPAIPGYENHGY.....VYGNVMVFTDIYKQMRADG 492

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/sgn2_1/USPTO_spool/US09615285/tunat_01062004_132133_19962/app_query.fasta_1.647
-DB=EST -OFMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09615285 @CGN 1 1 3437 @tunat_01062004_132133_19962 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2674	98.4	1749	11	BC035623	BC035623 Homo sapi
2	2291	84.3	1452	29	AY419481	AY419481 Homo sapi
3	1770	65.1	1446	29	AY419483	AY419483 Mus muscu
4	1770	65.1	1452	29	AY419482	AY419482 Pan trog
5	1237.5	45.5	866	12	BI761763	BI761763 603046751
6	1198	44.1	781	13	BX508967	BX508967 DKF2P686C
7	1100	40.5	603	13	BX479625	BX479625 DKF2P686E
8	1095.5	40.3	921	12	BI414727	BI414727 602991112
9	1088	40.0	604	14	CD724002	CD724002 OJ29all.Y
10	1084	39.9	678	14	CF147359	CF147359 UI-HF-C80
11	1066	39.2	591	12	BM747154	BM747154 K-EST0082
12	1046.5	38.5	736	12	BI454294	BI454294 603171527
13	1040	38.3	767	14	C989588	C989588 AGENCOURT
14	1028	37.8	644	14	CF147031	CF147031 UI-HF-C80
15	1000	36.8	622	14	CF147378	CF147378 UI-HF-C80
16	961.5	35.4	697	12	BI454495	BI454495 603171455
17	914.5	33.7	796	12	BI556978	BI556978 603238439
18	914.5	33.7	944	13	BQ228388	BQ228388 AGENCOURT
19	895	32.9	1034	13	BQ241270	BQ241270 603322326
20	894	32.9	910	12	BG242090	BG242090 602354633
21	892.5	32.8	814	10	BF675496	BF675496 602138477
22	890.5	32.8	648	10	BF102443	BF102443 601751865
23	885	32.6	486	13	EX099916	EX099916 BX099916
24	873	32.1	480	12	BM771081	BM771081 K-EST0005
25	870.5	32.0	961	10	BF584569	BF584569 602098278
26	858	31.6	892	13	BUS24316	BUS24316 AGENCOURT
27	857.5	31.6	1137	12	BI687501	BI687501 603315055
28	844	31.1	737	10	BF168078	BF168078 601776336
29	838	30.8	529	9	AW058537	AW058537 wx23b08.x
30	833	30.7	592	13	BUS06558	BUS06558 603737747
31	830	30.5	816	13	BUS97449	BUS97449 603536894
32	823	30.3	639	12	BM784118	BM784118 K-EST0006
33	823	30.3	725	12	BM784096	BM784096 K-EST0006
34	813	29.9	528	10	BF076876	BF076876 226635.MA
35	813	29.9	838	12	BI144021	BI144021 602908124
36	806	29.7	465	10	B857738	B857738 7947f08.x
37	795	29.3	497	10	B857022	B857022 QV3-FT022
38	794	29.2	744	14	CB600520	CB600520 AGENCOURT
39	793	29.2	790	10	BF144549	BF144549 601790193
40	790	29.1	703	10	BE376601	BE376601 601226430
41	786	28.9	915	12	BI143513	BI143513 602907696
42	783	28.8	783	12	BG176274	BG176274 602338030
43	781.5	28.8	846	12	BG870706	BG870706 602791694
44	781	28.7	693	10	B8916896	B8916896 601666437
45	764	28.1	450	9	AI393270	AI393270 tq09d11.x

ALIGNMENTS

RESULT 1
BC035623
LOCUS
DEFINITION Homo sapiens, similar to kallikrein B, plasma (Fletcher factor) 1,
clone IMAGE:5187060, mRNA.
1749 bp mRNA linear HTC 20-SEP-2002
ACCESSION BC035623.1 GI:23242886
VERSION BC035623
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1749)

Strausberg, R.
 Direct Submission
 Submitted (31-JUL-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: gcabs-x@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nhgri.nih.gov
 Akhtar, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granits, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Teague, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 68 Row: 1 Column: 6
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 14602458
 This clone has the following problem: frame shifted.
 FEATURES
 Location/Qualifiers
 1..1749
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5187060"
 /tissue types="Colon, kidney, stomach, adult, whole pooled"
 /clone_1in="NIH MGC_11c"
 /lab_host="DH10B"
 /note="Vector: pOMV-SPORT6"
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,19e-241 Length: 1749
 Score: 2674.00 Matches: 488
 Percent Similarity: 99.19% Conservative: 1
 Best Local Similarity: 98.99% Mismatches: 1
 Query Match: 98.42% Indels: 3
 DB: 11 Gaps: 1
 US-09-615-285B-2 (1-492) x BC035623 (1-1749)
 QY 1 MetAlaLeuAnSerGlySerProProAlaIleGlyProTyrTyrGluAnHisGlyTyr 20
 DB 117 ATGGCTTTGAACCTCAGGGTACCACCGAGCTATTGGACCTTACTATGAAACCCATGGATAC 176
 QY 21 GlnProGluAnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 DB 177 CAACCGGAAACCCCTATCCCGCACACGCCACCTGTGTGCCCTGCTACGAGGTGCAT 236
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 DB 237 CCGGCTCAGTACTACCGTGTCCCGTGTCCCGAGTACGCCCGCGGGTCTCTGACGAGGCT 296
 QY 61 SerAnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 DB 297 TCCAAACCCCGTGTGTGACGACGCCCAATCCCATCCGGGACAGTGTGCACCTCAAG 356
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 DB 357 ACTAAGAAAGCACGTGCATCACCTTGACCTCGGGGACCTTCTCTGTGGGTGTAGGGGATACAAG 1490
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAnSerGlyIleGluCys 120
 DB 417 GCCGCTGGCCCTACTCTGGAGTTTCATGGCAGCAAGTGTCCCACTCTGGGATAGAGTGC 476
 QY 121 AspSerSerGlyThrCysIleAnProSerAnTyrCysAspGlyValSerHisCysPro 140
 DB 477 GACTCTCAGTACTCTGATCAACCCCTCTAATCTGGTGTATGGCGTGTACACTGCCCC 536
 QY 141 GlyGlyGluAspGluAnArgCysValArgLeuTyrGlyProAnPheIleLeuGlnVal 160
 DB 537 GCGCGGAGGACGAGAAATCGGTGTGTTCGCTCTACGGACCAAACTTCTATCTTCAGATG 596
 QY 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspTyrAnGluAnTyr 180
 DB 597 TACTCATCTCAGAGGAAGTCTGGCACCCCTGTGGCAAGACGACTGGAGACGAGACTAC 656
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAnAnPheTyrSerSerGlnGlyIle 200
 DB 657 GCGCGGCGGCTCGCAGGACATGGCTATTAAGAAATAATTTTACTCTAGCAAGGAATA 716
 QY 201 ValAspSerGlySerThrSerPheMetLysLeuAnThrSerAlaGlyAsnValAsp 220
 DB 717 GTGGATGACAGCGGATCCACGAGCTTTATGAATCTGAACCAAGTGGCGGCAATGTGCAT 776
 QY 221 IleTyr-LysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuAr 240
 DB 777 ATCTATAAAAC-----CACAGTGTGCTGTCTTCAAAGCAGTGGTTTCTTACG 830
 QY 240 GCysIleAlaCysGlyValAnLeuAnSerSerArgGlnSerArgIleValGlyGly 260
 DB 831 CTGTATAGCTTCGCGGGTCAACTTGAATCAAGCCGCCAGAGCAGGATCGTGGCGGTGA 890
 QY 260 uSerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAnValHisVa 280
 DB 891 GAGCGCGCTCCCGGGGCGCTGGCCCTGGCAGGTGAGCTGCAGCTCCAGACGCTCCAGT 950
 QY 280 lCysGlyGlySerIleIleThrProGluTyrPileValThrAlaAlaHisCysValGly 300
 DB 951 GTGGGAGGCTCCATCATCACCCCGAGTGTGATGTGCAGCCGCCCTCATCGCTGGGAAA 1010
 QY 300 sProLeuAnAnProTyrHisThrAlaPheAlaGlyIleLeuArgGlnSerPheMe 320
 DB 1011 ACCTCTTAACAATCCATGGCATTTGACCGCATTTGCGGGATTTTGACACATCTTTCAT 1070
 QY 320 tPheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAnTyrAspSerLy 340
 DB 1071 GTTCTATGGAGCGGATACCAAGTAGAAAAAGTAGATTTCTCATCAAATTTATGACTCCA 1130
 QY 340 sThrLysAnAnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAnAspLe 360
 DB 1131 GACCAAGAACAAATGACATTCGCTGATGAAGCTGCAAGAGCTCTGACTTTCAAGACCT 1190
 QY 360 uValLysProValCysLeuProAnProGlyMetMetLeuGlnProGluGlnLeuCysTr 380
 DB 1191 AGTGAACACAGTGTGTGCCCCAACCCAGGCGATGATGTGCAGCCAGAACAGCTCTGTGT 1250
 QY 380 pLysSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAnAlaAl 400
 DB 1251 GATTTCCGGGTGGGGGCGCCACCGAGGAGAAAGGAAAGACCTCAGAAGTGTGAAGCTGC 1310
 QY 400 aLysValLeuLeuLeuGluThrGlnArgCysAnSerArgTyrValTyrAspAnLeuI 420
 DB 1311 CAAGGTGTTTCTATTGAGACACAGAGATGCAACAGCAGATATGTCTATGACACCTGAT 1370
 QY 420 eThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAnValAspSerCysGlnGlyAs 440
 DB 1371 CACACCGCATGATCTGTGCGGCTTCTGTGAGGAACTCGATTCCTTCCAGGGGGA 1430
 QY 440 pSerGlyGlyProLeuValThrSerLysAnAnIleTyrTrpLeuIleGlyAspThrSe 460
 DB 1431 CAGTGGAGGGCTCTGGTCACTTCGAGAAACAATATCTGTGGTGTGTAGGGGATACAAG 1490

```

QY 460 rTgGlySerGlyCysAlaValArgProGlyValTyrGlyAsnValMetValph 480
Db 1491 CTGGGGTCTGGCTGCGCAAGCTTACAGACGAGGTGACGGAGATGATGATTT 1550
QY 480 eThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1551 CACGAGCTGGATTATCGACAAATGAGGGCAGACGGC 1587

RESULT 2
AY419481 1452 bp DNA linear GSS 17-DEC-2003
LOCUS Homo sapiens TMPSR2 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY419481
VERSION AY419481.1 GI:39775438
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1452)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1452)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1452
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1452
/gene="TMPSR2"
/locus_tag="HCM6905"
Gene
ORIGIN
Alignment Scores: 4.44e-205 Length: 1452
Pred. No.: 2291.00 Matches: 424
Score: 87.60% Conservative: 0
Percent Similarity: 87.60% Mismatches: 60
Best Local Similarity: 84.32% Indels: 0
Query Match: 29 Gaps: 0
DB: 0

US-09-615-285B-2 (1-492) x AY419481 (1-1452)
QY 6 GlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyrGlnProGluAsnPro 25
Db 1 GGGTCAACCAAGCTATTGGACCTTACTATGAAACCAATGGATACCAACCGGAAACCCC 60
QY 26 TyrProAlaGlnProThrValValProThrValTyrGluValHisProAlaGlnTyrTyr 45
Db 61 TATCCGCAACAGCCCACTGTGTGCCACTGTCTACGAGGTGCATCCGCTCAGTACAC 120
QY 46 ProSerProValProGlnTyrAlaProArgValLeuThrGlnAlaSerAsnProValVal 65
Db 121 CCGTCCCCCGTCCCGACGACCGCCCGAGGTCCTGACGACGGCTCCCAACCCCGCTGTC 180
QY 66 CysThrGlnProLysSerProSerGlyThrValCysThrSerLysThrLysLysAlaLeu 85

```

```

181 TGCACGCAGCCCAAAATCCCCATCCGGGACAGTGTGCACCTCAAGACTAAGAAAGCACTG 240
QY 86 CysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeuAlaGlyLeuLeu 105
Db 241 TGCATACCTTGACCTCGGGACCTTCTCTGGAGCTCGCTGGCGCTGCTGCTACTC 300
QY 106 TrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCysAspSerSerGlyThr 125
Db 301 TGGAGTTTCATGGGACGAGTGTCTCAACTCTGGGATAGAGTGCAGTCTCTCAGGTACC 360
QY 126 CysIleAsnProSerAsnTrpCysAspGlyValSerHisCysProGlyGlyGluAspGlu 145
Db 361 TGCATCAACCCCTTAACCTGTGTGATGGCTGTACACTGCCCGCGGGAGGAGCAG 420
QY 146 AsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnValTyrSerSerGlnArg 165
Db 421 AATCGGTGTCTTGGCTCTAGGACCAAACTTCTCATCTTCTCAGGTACTCATCTCAGAGG 480
QY 166 LysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyrGlyArgAlaAlaCys 185
Db 481 AAGTCTCTGGACCCCTGTGTGCCAAGACGACTGGAACGAGAACTACGGCGCGGGCGCTGC 540
QY 186 ArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIleValAspAspSerGly 205
Db 541 AGGACATGGCTATAGAATAATTTTACTCTAGCCAGGAATAGTGGATGACAGCGGA 600
QY 206 SerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLysLysLeu 225
Db 601 TCCACCAAGCTTTATGAACTGAACAAAGTCCGCAATGTCGATATCTATAAAAACTG 660
QY 226 TyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCysGly 245
Db 661 TACCACAGTGTATGCTCTTCTCAAAGACAGTGTCTTACGCTATAGCTGTGCGGG 720
QY 246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGly 265
Db 721 GTCAACTTGAACCTCAAGCGCCAGAGCAGGATTGTGGCGCGAGAGCGCTCCCGGG 780
QY 266 AlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
Db 781 GCCTGGCCCTGGCAGGTCAAGCTGCAGTCCAGAACGTCACAGTGTGGAGGCTCCATC 840
QY 286 IleThrProGluTrpIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnPro 305
Db 841 ATCAACCCCGAGTGGATCTGTGACAGCGCCCACTGCGTGGAAANNNNNNNNNNNNN 900
QY 306 TrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAlaGly 325
Db 901 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 960
QY 326 TyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAsnAsp 345
Db 961 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
QY 346 IleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCys 365
Db 1021 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1080
QY 366 LeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGlyTyrGly 385
Db 1081 CTGCCCCAACCCAGCATGATGCTGACGACGAGCAGCTCTGCTGGATTTCGGGTGGGG 1140
QY 386 AlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuIle 405
Db 1141 GCCACCGGAGAAAGGAGACCTCAGAAAGTGTGAACGCTGCAAGGTGTCTTCATT 1200
QY 406 GluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIle 425
Db 1201 GAGACAGAGATCAACAGCAGATATGTCTATGACAACTGATCACACGACCATGATC 1260
QY 426 CysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyProLeu 445
Db 1261 TGTCCGGCTTCTTGCAGGGGAAAGTGTGATTCTTGTCCAGGGTGACAGTGGAGGCTCTG 1320

```

```

446 ValThrSerLysAsnAsnIleTrrProLeuIleGlyAspThrSerTrpGlySerGlyCys 465
1321 GTCACTTCGAGACAAATATCTGGTGGCTGATAGGGGATACAGCTGGGGTCTGGCTGT 1380
466 AlaLysAlaLysThrArgProGlyValTyrGlyAsnValMetValPheThrAspTrrIleTyr 485
1381 GCCAAGCTTACAGACAGAGGTGTCAGCGGAATGTGGTATTACGAGACTGGATTAT 1440
486 ArgGlnMetArg 489
1441 CGACAAATGAGG 1452

RESULT 3
LOCUS AY419483
DEFINITION Mus musculus TMPRSS2 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY419483
VERSION 1
KEYWORDS GSS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1446)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tannenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Perrier,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1446)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tannenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Perrier,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Source 1..1446
/mol_type="Mus musculus"
/db_xref="taxon:10090"
<1..>1446
/gene="TMPRSS2"
/locus_tag="HCM6905"

ORIGIN
Alignment Scores:
Pred. No.: 6,766-156 Length: 1446
Score: 1770.00 Matches: 330
Percent Similarity: 75.41% Conservative: 35
Best Local Similarity: 68.18% Mismatches: 117
Query Match: 65.15% Indels: 2
DB: 29 Gaps: 2

US-09-615-285B-2 (1-492) x AY419483 (1-1446)

QY 6 GlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyrGlnProGluAsnPro 25
DB 1 GGGTCACCTCCAGGAATCGACCTTGCTATGAGAACACCGGCTATCAGTCTGAGCACATC 60
QY 26 TyrProAlaGlnProThrValValProThrValTyrGluValHisProAlaGlnTyrTyr 45
DB 61 TGTCTCCGAGACACCGAGTGGCTCCCAATGGCTACACTGTATGATCCAGCCCATCTAC 120
QY 46 ProSerProValProGlnTyrAlaProArgValLeuThrGlnAlaSerAsnProValVal 65

```

```

121 CCAATCTCCAGTGCCTCAGATATGCTCCGAGGATTTACACGCAAGCCTCAACATCTGTCTC 180
QY 66 CysThrGlnProLysSerProSerGlyThrValCysThrSerLysThrLysLysAlaLeu 85
191 CACACACATCCCAAGTCC---TCAGGAGCACCCGTCGACCTCAAAGTCTAAGAAATCCTG 237
QY 86 CysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeuAlaAlaGlyLeuLeu 105
238 TGTTTAGCCCTTGCCCTTGCGCACTGCTCTCACGGAGAGTGTGTGGTGTCTGTCTTCTT 297
QY 106 TrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCysAspSerSerGlyThr 125
298 TGGAGGTTCGGGACAGCACTGTTCTACGTTCTGAGATGGAGTGTGGTCTTCAGGCACA 357
QY 126 CysIleAsnProSerAsnTrpCysAspGlyValSerHisCysProGlyGlyGluAspGlu 145
358 TGCATCAGCTCTCTCTCTGTGTGTGACGGGTGACACATTTGTCCTCCAAACGGAAGATGAG 417
QY 146 AsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnValTyrSerSerGlnArg 165
418 AACCGTTGTGTGTGTTCTACGGACAAAGCTTCATCTCTCAGGTTTACTCTTCAGAGG 477
QY 166 LysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyrGlyArgAlaAlaCys 185
478 AAAGCTCGTATCCCGTGTGCCAGGATGATTGGAGTCAGAGCTACGGGACAGCAGCATGT 537
QY 186 ArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIleValAspSerGly 205
538 AAAGACATGGGATCAAGAACAAATTTTATTCTAGCCAAAGGATACACAGACGAGCGGG 597
QY 206 SerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLysLysLeu 225
598 GCACAGAGCTTTATGAGCTGAATGTGAGCTCAGGCAATGTGACCTCTATAAAACTC 657
QY 226 TyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCysGly 245
658 TACCACAGTACTGATGTTATCCCGCATGGTGTTCTTTTGCGCTGTATAGAAATCGGG 717
QY 246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGly 265
718 GTTCGC---TCAGTGAAACGCCAGCAGCAGGATTTGGGTGGATTGAATGCTCACCAGGA 774
QY 266 AlaTrpProTrrGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
775 GACTGGCCCTGGCAGGTCAGCTGCACGTCAAGGCGTCCACGTCTCGGAGGCTCCATC 834
QY 286 IleThrProGluTrrIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnPro 305
835 ATCACCCCGAGTGGATTGTGCGCGCCCACTGTGTGGAAGANNNNNNNNNNNNNNNN 894
QY 306 TrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAlaGly 325
895 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 954
QY 326 TyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAsnAsnAsp 345
955 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1014
QY 346 IleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCys 365
1015 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1074
QY 366 LeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGlyTrrGly 385
1075 CTGCGGAACCCAGGCATGATGCTAGACCTAGACCTAGACCTGCTGGATTCTCGGGTGGGG 1134
QY 386 AlaThrGluGlyLysGlyThrSerGluValLeuAsnAlaAlaLysValLeuLeuIle 405
1135 GCCACCTATGAGAAGGAAAGACCTCGAGCTGTGAATGCTGCCATGTGTACCTTGATC 1194
QY 406 GluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIle 425

```

```

Db      1195 GAGCCCTCCAAATGTAATAGTAATAATACATACACCAACCTAATACACACGCGCATGATC 1254
Qy      426 CysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGlyProLeu 445
Db      1255 TGTCCGCGCTTCTCCAGGGGTCTGTGCACCTTTTCCAGGGAGACATGGAGGGCCCTG 1314
Qy      446 ValThrSerLysAsnAsnIleTrpPheLeuGlyAspThrSerTrpGlySerGlyCys 465
Db      1315 GTTACTTTGAAGAATGGGATCTGTGGTGTGATTGGGACACAGCTGGGGCTCGGGCTGT 1374
Qy      466 AlalaValAlaTyArgProGlyValTyGlyAsnValMetValPheThrAspTrpIleTy 485
Db      1375 GCCAAGGCACATCAGACCTGGAGTATACGGGAACGTGACGGTATTACAGATTGATCTAC 1434
Qy      486 ArgGlnMetArg 489
Db      1435 CAGCAAAATGAGG 1446

RESULT 4
AY419482
LOCUS
DEFINITION
  Pan troglodytes TMPRSS2 gene, VIRTUAL TRANSCRIPT, partial sequence,
  genomic survey sequence.
ACCESSION
  AY419482
VERSION
  AY419482.1 GI:39775439
KEYWORDS
  GSS.
SOURCE
  Pan troglodytes (chimpanzee)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
  1 (bases 1 to 1452)
AUTHORS
  Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
  Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
  Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
  Adams,M.D. and Cargill,M.
TITLE
  Inferring nonneutral evolution from human-chimp-mouse orthologous
  gene trios
JOURNAL
  Science 302 (5652), 1960-1963 (2003)
PUBMED
  14671302
REFERENCE
  2 (bases 1 to 1452)
  Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
  Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
  Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
  Adams,M.D. and Cargill,M.
DIRECT SUBMISSION
  Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
  Rockville, MD 20850, USA
COMMENT
  These sequences were made by sequencing genomic exons and ordering
  them based on alignment.
FEATURES
  source
    1..1452
    /organism="Pan troglodytes"
    /mol_type="genomic DNA"
    /db_xref="taxon:9598"
    <1..>1452
    /gene="TMPRSS2"
    /locus_tag="HCM6905"
ORIGIN

Alignment Scores:
Pred. No.:      6,8e-156      Length:      1452
Score:          1770.00      Matches:      337
Percent Similarity: 70.04%      Conservative: 2
Best Local Similarity: 69.63%      Mismatches: 145
Query Match:     65.15%      Indels:      0
DB:              29          Gaps:      0

US-09-615-285B-2 (1-492) x AY419482 (1-1452)
Qy      6 GlySerProAlaIleGlyProTyTrpGluAsnHisGlyTyGlnProGluAsnPro 25
Db      1 GGGTCACCAACGCTATTGGACCTTACTATGAAACCATGGATACCAACCGGAAACCCC 60

```

```

Qy      26 TyrProAlaGlnProThrValValProThrValTyGluValHisProAlaGlnTyTrp 45
Db      61 TATCCCGCACACGCCCACTGTGGCCCCCACTGTCTACGAGGTGCATCCGCTCAGTACTAC 120
Qy      46 ProSerProValProGlnTyAlaProArgValLeuThrGlnAlaSerAsnProValVal 65
Db      121 CGGTCCCCCGTGGCCAGTACGCCCGGAGGTCCTGACGACAGCTTCCAAACCCCGTGGTC 180
Qy      66 CysThrGlnProLysSerProSerGlyThrValCysThrSerLysThrLysLysAlaLeu 85
Db      181 CGCATGCAGCCCAAAATCCCACTCCGGGACAGTGTGCACCTCAANNNNNNNNNNNNNNNN 240
Qy      86 CysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeuAlaGlyLeuLeu 105
Db      241 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300
Qy      106 TrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCysAspSerSerGlyThr 125
Db      301 NNNNNNNNNNTGGCAGCAAGTCTCCAACCTGGGATAGATGGGACTCTCTCAGGTACC 360
Qy      126 CysIleAsnProSerAsnTrpCysAspGlyValSerHisCysProGlyGlyGluAspGlu 145
Db      361 TGCATCAGCCCTCTTAACCTGGTGTGATGGCGTGCACCTGCCCCACGCGGGAGACGAG 420
Qy      146 AsnArgCysValArgLeuTyGlyProAsnPheIleLeuGlnValTyTrpSerGlnArg 165
Db      421 AATCGGTGTGTTCGCTCTACGGACCAACTTCATCTTCAGTGTACTCATCTCAGAGG 480
Qy      166 LysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyGlyArgAlaAlaCys 185
Db      481 AAGTCTCGCACCTGTGTGCCAAGACGACGTGGAACGAGACTACGGCGGGCGGCTCG 540
Qy      186 ArgAspMetGlyTyTrpLysAsnAsnPheTyTrpSerGlnGlyIleValAspSerSerGly 205
Db      541 AGGACATGGGCTATGAAGATATATTTTACTTAGCCAAAGGATATAGTGGATGACACGGA 600
Qy      206 SerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyTrpLysLeu 225
Db      601 TCCACACAGCTTTATGAACACTGAACAAGTGGCGCATGTGATATCTATAAAAAACTG 660
Qy      226 TyrHisSerAspAlaCysSerSerLysAlaValSerLeuArgCysIleAlaCysGly 245
Db      661 TACCACAGTATGCCCTGTCTTCAAAGCAGTGGTTCTTTACGCTGTATAGCTCGGG 720
Qy      246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGly 265
Db      721 GTCAACTTGAACCTCAAGCCGCCACAGACGAGATCGTGGTGGCGAGAGCGCTCCCGGG 780
Qy      266 AlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
Db      781 GCCTGGCCCTGGCAGGTACGCTGCACGTCCAGAACGTCCAGTGTGCGGAGGCTCAATC 840
Qy      286 IleThrProGluTrpIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnPro 305
Db      841 ATCACCCCGCAGTGGATCGTGACAGCGCCCACTGGCTGGTGGAAANNNNNNNNNNNN 900
Qy      306 TrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyGlyAlaGly 325
Db      901 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 960
Qy      326 TyrGlnValGluLysValIleSerHisProAsnTyArgSerLysThrLysAsnAsnAsp 345
Db      961 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
Qy      346 IleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCys 365
Db      1021 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1080
Qy      366 LeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGlyTrpGly 385
Db      1081 CTGCCCCAACCCAGCATGATGCTGGAGCCAGACAGCTCTGCTGATTTCCGGGTGGGG 1140
Qy      386 AlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuLeuIle 405

```


AUTHORS Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R.,
Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and
Wiemann, S.
TITLE EST (Poustka, A., Albert, R., Moosmayer, P., Schupp, I.,
Wellenreuther, R., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS

MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.

No s1 sequence available.
This clone (DKFZp686C09277) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Location/Qualifiers
1..781
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686C09277"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfIIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Alignment Scores:
Pred. No.: 2,466-102 Length: 781
Score: 1198.00 Matches: 215
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.54% Mismatches: 0
Query Match: 44.0% Indels: 0
DB: 13 Gaps: 0

US-09-615-285B-2 (1-492) x BX508967 (1-781)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 132 ATGGCTTTGAATCAGGCTCACCACCGACTATGGACCTTACTATGAAACCCATGGATAC 191
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
DB 192 CAACCGAAACCCCTATCCCGACAGCCCACTGTGGTCCCACTGTCTACGAGTGCAT 251
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 252 CCGGCTCAGTACTACCGCTGCCCGGTCCTCCAGTACGCCCCGAGGCTCTGACGAGGCT 311
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 312 TCCACCCCGTGTCTGTCAGCGAGCCCAATCCCATCCGACAGTGTGCACCTCAAAG 371
QY 81 ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
DB 372 ACTAAGAAGCACTGTGTCATCACTTGACCTGGGACCTTCTCTGGGAGCTCGGTG 431
QY 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 432 CGCGCTGGCTACTCTGGAAGTTCATGGGACAGCAAGTGTCCCACTCTGGGATAGATGC 491
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 492 GACTCCTCAGGTACCTGCATCAACCCCTCTAATGCTGTGATGGGTGTGCACACGCCCC 551
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 552 GCGCGGAGGACAGAAATCGGTGTGTTCGCTCTACGGACCAACTTCTATCTCTCAGATG 611

QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGlnuTyr 180
DB 612 TACTCATCTCAGAGGAAGTCTCTGGCACCCCTGTGTGCCAAGACGACTGGACGAACTAC 671
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 672 GGCAGCGCGCGCTCGCAGGACATGGCTATAGATAATTTTACTCTAGCCAGGAATA 731
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAla 216
DB 732 GTGGATGACAGCGGATCCACCGACTTATGAAACTGACACCAAGTGCC 779

RESULT 7

BX479625

LOCUS

DEFINITION

DKFZp686E22213_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone

DKFZp686E22213_5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by GBF (National Research Centre for Biotechnology Ltd.,

Braunschweig/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No s1 sequence available.

This clone (DKFZp686E22213) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Location/Qualifiers

1..603

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp686E22213"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="686 (synonym: hlcc3)"

/note="Vector: pTriplex2; Site_1: SfIIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Alignment Scores:

Pred. No.: 2,886-93 Length: 603

Score: 1100.00 Matches: 200

Percent Similarity: 99.50% Conservative: 0

Best Local Similarity: 99.50% Mismatches: 1

Query Match: 40.49% Indels: 0

DB: 13 Gaps: 0

US-09-615-285B-2 (1-492) x BX479625 (1-603)

QY 145 GluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnValTyrSerSerGln 164
DB 1 GAGAAATCGGTGTGTTCGCTCTACGGACTAACTTATCTCTCAGGTGTACTCTCTCAG 60
QY 165 ArgLysSerTrpHisProValCysGlnAspAspTrpAsnGlnuTyrGlyArgAlaAla 184
DB 61 AGGAAGTCTCTGGCACCCCTGTGTGCCAAGACGACTGGACGAACTACGGCGCGCGGCC 120


```

185 CysArgAspMetGlyTyrLeuAsnAsnProPheTyrSerSerGlnGlyValLeuAspAspSer 204
121 TGCAGGAGACATGGCTATAGATAATATTTTACTCTAGCCAAAGGAATAGTGATGACAGC 180
205 GlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLysLys 224
181 GGATCCACCAAGCTTTATGAACCTGAACCAAGTGGCGCAATGTCTGATATCTATAAAAA 240
225 LeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCys 244
241 CTGTACACAGATGATGCTGTTCTTCAAGACAGTGTGTTCTTTAGCTGTATAGCTTGC 300
245 GlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlySerAlaLeuPro 264
301 GGGGTCAACTGAACCTCAAGCCCGCAGACGAGGATTTGGGGCGGCGAGCGCTCCCG 360
265 GlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySer 284
361 GGGGCTGGCCCTGGCAGGTACCTTGCAGTCCAGACGTCACAGTGTGGAGGCTCC 420
285 IleIleThrProGlnTyrIleValThrAlaAlaHisCysValGlyLysProLeuAsnAsn 304
421 ATCATCAACCCCGAGTGGATCGTGACAGCGCCCACTGGGTGGAAAAACCTCTTAACAA 480
305 ProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAla 324
481 CCATGGCATTTGGACGGCAATTTGGGGGATTTTGAGACAATCTTTCAATGTTCTATGG 540
325 GlyTyrGlnValGlnLysValIleSerHisProAsnTyrAspSerLysThrLysAsnAsn 344
541 GGATACCAAGTAGAAAAAGTATTTCTCTCAAAATTATGACTCCAAAGACCAAGAACAA 600
345 Asp 345
601 GAC 603

RESULT 8
BI414727
LOCUS
DEFINITION
602991112F1 NC1_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5147162 5',
mRNA sequence.
ACCESSION
BI414727
VERSION
BI414727.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 921)
NTH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: NC1-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11363 row: 9 column: 03
High quality sequence start: 29
High quality sequence stop: 873.
FEATURES
Location/Qualifiers
1..921
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:5147162"

```

```

/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NC1_CGAP_Lu33"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site: 1: NotI; Site: 2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCTCTGTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

```

ORIGIN

Alignment Scores:

```

Pred. No.: 1-53e-92 Length: 921
Score: 1095.50 Matches: 206
Percent Similarity: 82.93% Conservative: 32
Best Local Similarity: 71.78% Mismatches: 47
Query Match: 40.32% Indels: 3
DB: 12 Gaps: 1

```

US-09-615-285B-2 (1-492) x BI414727 (1-921)

```

QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 65 ATGGCATTTGAATCAGGCTCACCCTCCAGGATCCGACTTGCATATGAGACCCGGGTAT 124
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 125 CAGTCTGAGCACATCTGTCTCCGAGACACACAGTGGCTCCCAATGGCTCAACTTCTGAT 184
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 185 CCAGCCAGTACTACCCATCTCCAGTGCTCAGTATGCTCCGAGATTACACGCAAGCC 244
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 245 TCAACATCTGTCTATCCACACACATCCCAAGTCC---TCAGGAGCACCGCTGCACCTCAAG 301
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 302 TCTAGAAATCGCTGTGTAGCCCTTGGCTGGCACTGTCTCCACGGGAGCTGCTGTG 361
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 362 GTGTGTCTGTCTTGGAGGTTCTGGACAGCAACTGTTCTACGCTCTGAGATGGAGTGT 421
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
DB 422 GGGTCTTCAGGCACATGATCAGCTCTCTCTCTGGTGTGACGGGTAGCACATTGTGCC 481
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 482 AACGGAGAGATGAGACCGTGTGTCTCTACGGACAAAGCTTCATCCCTCCAGGTT 541
QY 161 TyrSerSerGlnArgLysSerTyrPheIleProValCysGlnAspAspTyrAsnGluAsnTyr 180
DB 542 TACTCATCTCAGAGGAAAGCCTGTATCCCGTGTGCGCAGGATGATGAGTAGAGAGTAC 601
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 602 GGGAGAGCAGCATGTAAAGACATGGATACAGAACCAATTTTATTCTACCAAGGAGATA 661
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 662 CCAGACCAAGCGGGCAACGAGCTTTATGAAGCTGAATGTGAGCTCAGGCAATGTTGAC 721
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 722 CTCTATAAAACACTCTACCAAGTGAATGACTTCAGTTTCATCCCGCATGTTGTTCTTTGCGC 781

```


/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30571118"
 /tissue_type="CNCAP (3) T-225 cell line"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH MGC 210"
 /note="Organ: Prostate; Vector: pRTT3 Pac; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pRTT3 Pac vector. The library tag sequence located between the Not I site and the polyA tail is CCCAC. Tissue was provided by Tim Ratliff."

ORIGIN

Alignment Scores:

Pred. No.: 1.13e-91 Length: 678
 Score: 1084.00 Matches: 196
 Percent Similarity: 99.49% Conservative: 0
 Best Local Similarity: 99.49% Mismatches: 1
 Query Match: 39.90% Indels: 1
 DB: 14 Gaps: 0

US-09-615-285B-2 (1-492) x CF147359 (1-678)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 88 ATGGCTTTGAACCTCAGGTCACACAGCATTATGGACCTTACTATGAACCATGGATAC 147
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 148 CAACCGGAAGAACCCCTATCCCGACAGCCACTGGTGGTCCCATCTGTCTACGAGGTGCAT 207
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 208 CCGGCTCAGTACTACCCGTCCTCCGTCAGTACGCCAGTCCGAGGCTCTGACGACGCT 267
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 268 TCACACCCCGTCTGTCAGCAGGCCCAATCCCATCCGAGCAGTGTGCACCTCAAG 327
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 328 ACTAAGAAAGCACTGTGCATCACTTACCTTGGGACCTTCTCGTGGAGCTGGCTG 387
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 388 GCGGCTGGGCTACTCTGGAAGTTCATGGGACCAAGTGTCTCAACTCTGGGATAGATGC 447
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 448 GACTCCTCAGGTACCTGCATCAACCCCTTAACCTGTGTGATGGGTGTGCACACTGCC 507
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 508 GCGGGGAGCGAGAAATCGGTGTCTCGCTCTACGGACCAAACTTCATCTCTCAAGT 567
 QY 161 TyrSerSerGlnArgLysSerTyrPheSerProValCysGlnAspAspTyrPheGluAsnTyr 180
 Db 568 TACTCATCTCAGAGAGTCTCTGGCACCCCTGTGTGCCAAGACGACTGGACGAGACTAC 627
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSer 197
 Db 628 GGGCGGGCGGCTCGAGGAGACATGGGCTATAAGATAT-ATTTACTCTAGC 677

RESULT 11

BM747154

LOCUS

591 bp mRNA linear EST 04-MAR-2002

DEFINITION

X-RST0021659 S6SNU620 Homo sapiens cDNA clone S6SNU620-20-F02 5', mRNA sequence.

ACCESSION BM747154

VERSION EM747154.1 GI:19076749

KEYWORDS

SOURCE EST

ORGANISM

Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.

TITLE

JOURNAL

COMMENT

21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Soeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 20 row: F column: 02
 High quality sequence stop: 591.
 Location/Qualifiers

FEATURES

source

1..591
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S6SNU620-20-F02"
 /sex="F"
 /tissue_type="Ascites"
 /cell_type="Scattering floating"
 /cell_line="SNU-820"
 /lab_host="Top10F"
 /clone_lib="S6SNU620"
 /note="Organ: Stomach; Vector: pCNS; Site 1: EcoRI; Site 2: NotI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

ORIGIN

Alignment Scores:

Pred. No.: 4.51e-90 Length: 591
 Score: 1066.00 Matches: 196
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 39.23% Indels: 0
 DB: 12 Gaps: 0

US-09-615-285B-2 (1-492) x BM747154 (1-591)

QY 224 LysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArgCysIleAla 243
 Db 2 AAACCTGACCAACAGTATGCTGTCTCTCAAGACAGTGGTCTTCTTACGCTATAGCC 61
 QY 244 CysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeu 263
 Db 62 TCGCGGGTCAACTTGAATCAAGCCGACAGCAGGATTGTGGCGCGCGAGCGGCTC 121
 QY 264 ProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisValCysGlyGly 283

```

Db 122 CCGGGGGCTGGCCCTGGCAGCTGCGCTCCAGAACCTCCACGCTGTGCGGAGGC 181
QY 284 SerileileThrProGluThrIleValThrAlaAlaHisCysValGluLysProLeuAen 303
Db 182 TCCATCATCACCCCGAGTGGATCGTGACAGCGCCGACCTGGTGGAAAACCTCTTAC 241
QY 304 AsnProThrHisThrAlaPhaAlaGlyIleLeuArgGlnSerPheMetPheTyrGly 323
Db 242 AATCATGGCAATGGACGGCAATTCGCGGGATTTGAGACAATCTTTTCATGTCTATGGA 301
QY 324 AlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAen 343
Db 302 GCGGATACCAAGTAGAAAAGTATTTCTCATCCAAATATGACTCCAGACCAAGAAC 361
QY 344 AsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysPro 363
Db 362 AATGACATTTGGCTGTGATGAAGCTGCAGAGCCCTCTGACTTTCAACGACTAGTGAACCA 421
QY 364 ValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGly 383
Db 422 GTGTGTCTGCCAACCCAGCATGATGCTGCAGCCAGACAGCTCTGCTGGATTTCCGGG 481
QY 384 TrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaLysValLeu 403
Db 482 TGGGGGGCCACCGAGGAGAAAGGAGAACCTCAGAAAGTGTGAACGCTGCCAAGGTGCTT 541
QY 404 LeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeu 419
Db 542 CTCATTGACACAGAGATGCACAGACAGATATGCTATGACACACCTG 589

```

RESULT 12
B1454294
LOCUS 603171527F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5250822 5',
DEFINITION mRNA sequence.

ACCESSION B1454294
VERSION B1454294.1 GI:15244950
KEYWORDS EST.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM11633 row: f column: 07
High quality sequence stop: 736.

FEATURES
Location/Qualifiers
1..736
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5250822"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam5"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; Salt;
Site: 2; Nctt; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,

RESULT 13
CB989588
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS

CB989588 767 bp mRNA linear EST 01-MAY-2003
AGENCOURT 13894812 NIH MGC_147 Homo sapiens cDNA clone
IMAGE:30347890 5', mRNA sequence.

CB989588
CB989588.1 GI:30284108
EST.

ORIGIN
Alignment Scores:
Pred. No.: 4,48e-88 Length: 736
Score: 1046.50 Matches: 193
Percent Similarity: 87.45% Conservative: 28
Best Local Similarity: 78.14% Mismatches: 23
Query Match: 38.52% Indels: 3
DB: 12 Gaps: 1
US-09-615-285B-2 (1-492) x B1454294 (1-736)
QY 98 AlaAlaLeuAlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGly 117
Db 2 GCTGCTGTGCTGCTCTGCTGCTTTGGAGGTCTTGGACAGCAACTGTTCTACGCTCTAG 61
QY 118 IleGluCysAspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSer 137
Db 62 ATGAGTGTGGTCTTCAGGCACATGCATCAGCTCTCTCTCTGCTGTGACGGGTAGCA 121
QY 138 HisCysProGlyGlyClnAspGluAsnArgCysValArgLeuTyrGlyProAsnPhelle 157
Db 122 CATGTGCCCAACGAGAGATGAGAACCTTGTGTCTCTACGGACAAAGCTTCATC 181
QY 158 LeuGlnValTyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsn 177
Db 182 CTCAGGTTTACTCATCTCAGAGGAAGCTGCTATCCGCTGTCAGAGTATTGGAGT 241
QY 178 GluAsnTyrGlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPhetyrSerSer 197
Db 242 GAGAGCTCGGGAGGAGCAGCATGTAAGACATGGGATACAGAACAAATTTTATTCTAGC 301
QY 198 GlnGlyIleValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGly 217
Db 302 CRAAGGATACACAGACCGGGCGGCAACGACCTTATGAGCTGAATGTGAGCTCAGGC 361
QY 218 AsnValAspIleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValVal 237
Db 362 AATGTTGACCTCTATAAAAACTCTACCACAGTACTCATGTTCACTCCGATGGTGTG 421
QY 238 SerLeuArgCysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleVal 257
Db 422 TCTTTGCGCTGTATAGATCGGGGTTCGC--TCAGTGAACACGCGCAGCAGGATTGTG 478
QY 258 GlyGlyGluSerAlaLeuProGlyAlaThrProTyrGlnValSerLeuHisValGlnAsn 277
Db 479 GGTGGATTGAATGCTCTACAGAGACTGGCCCTGCGAGGTGAGCTGCATCCCAAGGC 538
QY 278 ValHisValCysGlyGlySerIleIleThrProGluThrIleValThrAlaAlaHisCys 297
Db 539 GTCCACGCTCGGGAGGCTCCATCATCACCCCGAGTGGATTGTGACGGCGGCCACTGT 598
QY 298 ValGluLysProLeuAsnAspProTrpHis-TripThrAlaPheAlaGlyIleLeuArgG 317
Db 599 GTGGAAAGAACCCCTCAGCGGCCCGGAGGTACTGTGACGGCAATTCGGGAATCTGAGACA 658
QY 317 nSerPheMetPheTyrGlyAlaGlyTyrGlnValGluLys-ValIleSerHisProAsn 337
Db 659 GTCTCTCATGTTCTATGGAAGTAGACACAGGTAGAAACAGTAATTTCCCATCAAATT 718
QY 337 YrAspSerLysThrLys 342
Db 719 ACGACTCTAAGACCAAG 735

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 767)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Dr. Stefan Hansson
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
 and advice from Piero Carninci (RIKEN)
 CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDAM389 row: 1 column: 11
 High quality sequence stop: 676.

FEATURES
 source
 1..767
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30347890"
 /tissue_type="Human Placenta"
 /lab_host="DH10B Tona"
 /clone_lib="NIH_MGC_147"
 /note="Organ: Placenta; Vector: pBluescriptR; Site 1:
 all-XhoI; Site 2: BamH; Oligo-dr primed using primer
 5'-TTTTTTTTTTTTTTVN-3', size-selected for average
 insert size 2.3 kb and normalized to ROT 5. This is a
 primary library enriched for full-length clones and
 constructed using the Cap-trapper method (Carninci, in
 preparation). Library constructed by M. Brownstein, in
 (NHGRI/NHGRI, National Institutes of Health). Note: This is
 a NIH_MGC library."

ORIGIN

Alignment Scores:
 Pred. No.: 1,97e-87 Length: 767
 Score: 1040.00 Matches: 208
 Percent Similarity: 79.55% Conservative: 2
 Best Local Similarity: 78.79% Mismatches: 8
 Query Match: 38.28% Indels: 46
 DB: 14 Gaps: 4

US-09-615-285b-2 (1-492) x CB989588 (1-767)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 |||||
 DB 96 ATGGCTTTGAATCAGGTCACACAGCTATTGACCTTACTATGAAACCATGGATAC 155
 |||||
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 |||||
 DB 156 CAACCGGAAACCCCTATCCGACAGCCACTGTGGTCCCACTGTCTACGAGGTGCAT 215
 |||||
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 |||||
 DB 216 CCGGCTCATCTACTACCCGTCCTCCCGCTGCGCCAGTACGCGCCGAGGTCCTGACACAGCT 275
 |||||
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 |||||
 DB 276 TCCACACCCGCTGCTGTGACGAGCCCAATCCCATCCGAGCAGTGTGCACCTCAAG 335
 |||||
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 |||||
 DB 336 ACTAAGAAGAACTGTGCTATCCTTGACCTTGGGAGCTTCTCTGGGGAGCTGGCCTG 395
 |||||
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 |||||
 DB 396 GCGCTGGCTACTCTGGAAGTTC----- 419

QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 419 ----- 419
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 :|||
 DB 420 -----ATTGCGCTCTACGACCAAACTTCATCCCTTCAGGTG 455
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 |||||
 DB 456 TACTCATCTCAGAGGAGTCTCTGGACCTGTGCCAAGACGACTGAGACGAACTAC 515
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 DB 516 GGGCGGGGGCTGCGAGGACATGGCTATAGAAATATTTTACTCTAGCCAGGAATA 575
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 |||||
 DB 576 GTGGATGACAGGGGATCCACGCTTTATGAACTGACACAGTGGCGGCACTGTGCAT 635
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuAr 240
 DB 636 ATCTATAAAAAAAGTGTACACAGTGTCTTCTTCAAAAAGCAGTGGTTCCTTTTAC 695
 QY 240 GCysile---AlaCys-GlyValAsn---LeuAsnSerSer-----ArgGlnSerArgI 256
 |||||
 DB 696 NCTGTATAAGCTTGGGGGGTCCAACTTGAAGTCAAGCCCGCCAGGAGCCAGGAA 755
 QY 256 leValGly 258
 |||||
 DB 756 TCGTTGGG 763
 RESULT 14
 CF147031
 LOCUS
 DEFINITION UI-HF-C80-atg-a-09-0-UI-r1 NIH_MGC_210 Homo sapiens CDNA clone
 IMAGE:30570824 5', mRNA sequence.
 CF147031
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 644)
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Tim Ratliff
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/humanfl.html>
 Seq primer: pYX-5.
 FEATURES
 Location/Qualifiers
 1..644
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30570824"
 /tissue_type="CNCAP (3) T-225 cell line"

```
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_MGC_210"
/notes="Organ: Prostate; Vector: pT73 Pac; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with Not I and then cloned
directionally into pT73 Pac vector. The library tag
sequence located between the Not I site and the polyA tail
is CCCAC. Tissue was provided by Tim Ratliff."
```

ORIGIN

Alignment Scores:

Pred. No.:	2,01e-86	Length:	644
Score:	1029.00	Matches:	185
Percent Similarity:	99.46%	Conservative:	0
Best Local Similarity:	99.46%	Mismatches:	0
Query Match:	37.84%	Indels:	1
DB:	14	Gaps:	0

US-09-615-285B-2 (1-492) x CF147031 (1-644)

QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20

Db 88 ATGGCTTTGAACCTCAGGTCACACCCAGCTATTGGACCTTACTATGAAACCATGGATAC 147

QY 21 GlnPro-gluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHi 40

Db 148 CAACCGGAAACCCCTATCCCGACACGCCACCTGTGGTCCCCACTCTACGAGGTGCA 207

QY 40 sProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAl 60

Db 208 TCCGCTCAGTACTACCGCTCCCGCTGCCCGACGACGCGCGAGGTCCTGACGCGAGC 267

QY 60 aSerAsnProValValCysThrGlnProLysProSerGlyThrValCysThrSerLy 80

Db 268 TTCCAAACCCCGTGCCTCAGCGAGCCAAATCCCATCCGGAGAGTGTCACCTCAA 327

QY 80 sThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLe 100

Db 378 GACTAAGAAAGCAGCTGTGCATCACCTTGACCTGGGACCTCTCCTGGGAGCTGGCT 387

QY 100 uAlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCy 120

Db 388 GCGCGCTGGCTACTCTGGAAGTTTCATGGCGAGCAAGTCTCCAACTCTGGATAGAGTG 447

QY 120 aAspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPr 140

Db 448 CGACTCTCAGTACTCTGCATCAACCCCTCTAAGTGGTGTGATGGCGGTGCACACTGCC 507

QY 140 cGlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVa 160

Db 508 CGCGGGGAGGACGAGAGTCGGGTGTTCGCTCTACGACCAAACTTCATCCTTCAGGT 567

QY 160 lTyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTy 180

Db 568 GTACTCATCTCAGAGAGATCTCTGCACCCCTGTGTGCCAAGACGACTGGACGAGACTA 627

QY 180 rGlyArgAlaAlaCys 185

Db 628 CGGGCGGGCGGCTGC 643

RESULT 15

CF147378

LOCUS

DEFINITION

UI-HF-CB0-atj-d-03-0-UI.r1 NIH_MGC_210 Homo sapiens cDNA clone

IMAGE:30571142 5', mRNA sequence.

CF147378

ACCESSION

CF147378.1 GI:33262822

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 622)

Ronaldo,M.P., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Tim Ratliff

CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

CDNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/humanf1.html

Seq primer: pyx-5.

Location/Qualifiers

1..622

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:30571142"

/issue_type="NCAP(3)T-225 cell line"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH_MGC_210"

/note="Organ: Prostate; Vector: pT73 Pac; Site 1: EcoR I; Site 2: Not I; The library was constructed according

Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pT73 Pac vector. The library tag sequence located between the Not I site and the polyA tail is CCCAC. Tissue was provided by Tim Ratliff."

ORIGIN

Alignment Scores:

Pred. No.:	8,35e-84	Length:	622
Score:	1000.00	Matches:	178
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	36.81%	Indels:	0
DB:	14	Gaps:	0

US-09-615-285B-2 (1-492) x CF147378 (1-622)

QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20

Db 88 ATGGCTTTGAACCTCAGGTCACACCCAGCTATTGGACCTTACTATGAAACCATGGATAC 147

QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40

Db 148 CAACCGGAAACCCCTATCCCGACACGCCACCTGTGGTCCCCACTCTACGAGGTGCA 207

QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60

Db 208 CCGGCTCAGTACTACCGCTCCCGCTGCCCGACGACGCGCGAGGTCCTGACGCGAGCT 267

QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

Db 268 TCCAACCCCGTGTCTGACGACGAGCCAAATCCCATCCGGAGAGTGTCACCTCAAAG 327

```
QY      81 ThrLysLysAlaLeuCysIleThrLeuThrLeuThrLeuValGlyAlaAlaLeu 100
      |||
Db      328 ACTAAGAAAGCACTGTGCAATCCTTGACCTGGGACCTTCTCGTGGGAGCTGGCGTG 387
      |||
QY      101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
      |||
Db      388 GCCGCTGGGCTACTCTGGAGTTCATGGGCGACCAAGTCTCCAACTCTGGGATAGATGC 447
      |||
QY      121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
      |||
Db      448 GACTCCTCAGGTACCTGCAATCAACCCCTTAACCTGGTGTGATGGCGTGTCACTGCCCC 507
      |||
QY      141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
      |||
Db      508 GCGGGGGAGGACGAGATCGGTGTGTTCGCCCTTACGGACCAAACTTCATCTTCAGGTG 567
      |||
QY      161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGlu 178
      |||
Db      568 TACTCATCTCAGAGGAAGTCTGTGGCACCCCTGTGTGCCAAGACGACTGGAAACGAG 621
      |||
```

Search completed: June 1, 2004, 17:39:44
Job time : 3794 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:07:53 ; Search time 26 Seconds
(without alignments)
985.328 Million cell updates/sec

Title: US-09-615-285B-2

Perfect score: 2717

Sequence: 1 MALNSGSPPAIGYENHGY.....YGVNVMVFTDIYQMRADG 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2714	99.9	492	1 TMS2 HUMAN	O15393 homo sapien
2	2117	77.9	490	1 TMS2 MOUSE	O91088 mus musculus
3	880	32.4	453	1 TMS3 MOUSE	O81110 mus musculus
4	875.5	32.2	454	1 TMS3 HUMAN	P57727 homo sapien
5	680	25.0	435	1 TMS4 MOUSE	O8VCA5 mus musculus
6	676.5	24.9	437	1 TMS4 HUMAN	O9NR84 homo sapien
7	665.5	24.5	455	1 TMS5 MOUSE	P98072 bos taurus
8	660	24.3	1035	1 ENTX BOVIN	P98072 bos taurus
9	651.5	24.0	457	1 TMS5 HUMAN	O9B383 homo sapien
10	647	23.8	1019	1 ENTX HUMAN	P98073 homo sapien
11	638	23.5	1069	1 ENTX MOUSE	P97435 mus musculus
12	634	23.3	1034	1 ENTX PIG	P98074 sus scrofa
13	601	22.1	1042	1 COR1 HUMAN	O9Y5G5 homo sapien
14	577.5	21.3	1113	1 COR1 MOUSE	O92319 mus musculus
15	569.5	21.0	638	1 KAL HUMAN	P03952 homo sapien
16	568.5	20.9	855	1 ST14 MOUSE	P56677 mus musculus
17	565.5	20.8	417	1 HEP5 HUMAN	P05981 homo sapien
18	558.5	20.6	855	1 ST14 HUMAN	O9Y5Y6 homo sapien
19	556	20.5	638	1 KAL RAT	P14272 rattus norv
20	556	20.5	811	1 TMS6 MOUSE	O9DB10 mus musculus
21	547	20.1	436	1 HEP5 MOUSE	O35453 mus musculus
22	538	19.8	638	1 KAL MOUSE	P26262 mus musculus
23	536.5	19.7	811	1 TMS6 HUMAN	O81U80 homo sapien
24	533	19.6	625	1 FALL HUMAN	P03951 homo sapien
25	529.5	19.5	416	1 HEP5 RAT	O05511 rattus norv
26	527.5	19.4	418	1 HAPT HUMAN	O80235 homo sapien
27	517.5	19.0	812	1 PLMN MOUSE	P0918 mus musculus
28	504.5	18.6	324	1 TEST MOUSE	O9JH77 mus musculus
29	501	18.4	421	1 ACRO HUMAN	P10323 homo sapien
30	496	18.3	422	1 DES1 HUMAN	O9UL52 homo sapien
31	490.5	18.1	415	1 ACRO PIG	P08001 sus scrofa
32	489.5	18.0	436	1 ACRO MOUSE	P23578 mus musculus
33	486.5	17.9	761	1 NEUR MOUSE	O08762 mus musculus

ALIGNMENTS

RESULT 1

TMS2_HUMAN	STANDARD	PRT	492 AA.
ID	O15393; Q9BXX1;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Transmembrane protease, serine 2 precursor (EC 3.4.21.-).		
GN	TPRSS2 OR PRS10.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97468144; PubMed=9325052;		
RA	Paoloni-Giacobino A., Chen H., Peitsch M.C., Rossier C.,		
RA	Antonarakis S.E.;		
RT	"Cloning of the TPRS2 gene, which encodes a novel serine protease		
RT	with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.";		
RL	Genomics 44:309-320(1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21309059; PubMed=11414763;		
RA	Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;		
RT	"Mutation analyses of 288 candidate genes in human tumor cell lines.";		
RL	Genomics 74:352-364(2001).		
RN	[3]		
RP	SEQUENCE FROM N.A., AND MUTAGENESIS.		
RX	MEDLINE=21139112; PubMed=11245484;		
RA	Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Saffran D.C.,		
RA	Raitano A.B., Jakobovits A.;		
RT	"Catalytic cleavage of the androgen-regulated TPRS2 protease results		
RT	in its secretion by prostate and prostate cancer epithelia.";		
RL	Cancer Res. 61:1686-1692(2001).		
RN	[4]		
RP	TISSUE SPECIFICITY.		
RX	MEDLINE=21104370; PubMed=11169526;		
RA	Vaara M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;		
RT	"Expression of transmembrane serine protease TPRS2 in mouse and		
RT	human tissues.";		
RL	J. Pathol. 193:134-140(2001).		
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein. Activated by		
CC	cleavage and secreted.		
CC	-I- TISSUE SPECIFICITY: Expressed strongly in small intestine. Also		
CC	expressed in prostate, colon, stomach and salivary gland.		
CC	-I- SIMILARITY: Belongs to peptidase family S1.		
CC	-I- SIMILARITY: Contains 1 LDL-receptor class A domain.		
CC	-I- SIMILARITY: Contains 1 SRCR domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/)		

34	477	17.6	790	1	PLMN PIG	P06867 sus scrofa
35	475.5	17.5	431	1	ACRO RABIT	P45038 Oryctolagus
36	472.5	17.4	810	1	PLMN MACMU	P12545 macaca mulia
37	471	17.3	290	1	PR27 HUMAN	Q9BQR3 homo sapien
38	470.5	17.3	437	1	ACRO RAT	P29293 rattus norv
39	470.5	17.3	810	1	PLMN ERIEU	Q29485 erinaceus e
40	469.5	17.3	311	1	TRYG MOUSE	Q9QU17 mus musculus
41	467.5	17.2	321	1	TRYG HUMAN	Q9NR22 homo sapien
42	463	17.0	273	1	TRYT SHEEP	O9X8M2 ovis aries
43	462	17.0	317	1	BSS4 HUMAN	Q93ZM4 homo sapien
44	461	17.0	343	1	PS86 HUMAN	Q16651 homo sapien
45	460.5	16.9	306	1	BSS4 MOUSE	Q9ER10 mus musculus

or send an email to license@sib-sib.ch).

```

CC CC EMBL; U75329; AAC51784.1; -
DR DR EMBL; AF123453; RAD37117.1; -
DR DR EMBL; AF270487; AAK29280.1; -
DR DR HSP; P00763; IDPO.
DR DR MEROPS; S01.247; -.
DR DR Gnew; HGNC:11876; TMPRSS2.
DR DR MIM; 602060; -.
DR DR GO; GO:0005887; C: integral to plasma membrane; TAS.
DR DR GO; GO:0008236; F: serine-type peptidase activity; TAS.
DR DR InterPro; IPR009003; Cys Ser. trypsin.
DR DR InterPro; IPR002172; LDL receptor A.
DR DR InterPro; IPR001254; Peptidase S1.
DR DR InterPro; IPR001314; Peptidase S1A.
DR DR InterPro; IPR001190; Srcr receptor.
DR DR Pfam; PF00089; trypsin; 1.
DR DR PRINTS; PR00722; CHYMOTRYPSIN.
DR DR SMART; SM00192; LDLA; 1.
DR DR SMART; SM00202; SR; 1.
DR DR SMART; SM00200; TRYD; SPC; 1.
DR DR PROSITE; PS01209; LDLA; 1.
DR DR PROSITE; PS00888; LDLA; 2; 1.
DR DR PROSITE; PS00420; SRCR; 1; FALSE NEG.
DR DR PROSITE; PS0287; SRCR; 2; 1.
DR DR PROSITE; PS0240; TRYPSIN; 1.
DR DR PROSITE; PS00134; TRYPSIN; HIS; 1.
DR DR PROSITE; PS00135; TRYPSIN; SER; 1.
DR DR Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW Polymorphism. 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT CHAIN 256 492 CATALYTIC CHAIN.
FT FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT FT CATALYTIC CHAIN.
FT FT CYTOPLASMIC (POTENTIAL).
FT FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT FT EXTRACELLULAR (POTENTIAL).
FT FT LDL-RECEPTOR CLASS A.
FT FT SRCR.
FT FT SERINE PROTEASE.
FT FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT FT ACT_SITE 296 236 CHARGE RELAY SYSTEM.
FT FT ACT_SITE 345 345 CHARGE RELAY SYSTEM.
FT FT ACT_SITE 441 441 CHARGE RELAY SYSTEM.
FT FT SITE 255 256 CLEAVAGE (POTENTIAL).
FT FT DISULFID 113 126 BY SIMILARITY.
FT FT DISULFID 120 139 BY SIMILARITY.
FT FT DISULFID 133 148 BY SIMILARITY.
FT FT DISULFID 172 231 BY SIMILARITY.
FT FT DISULFID 185 241 BY SIMILARITY.
FT FT DISULFID 244 365 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 281 297 BY SIMILARITY.
FT FT DISULFID 410 426 BY SIMILARITY.
FT FT DISULFID 437 455 BY SIMILARITY.
FT FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT VARIANT 449 449 K -> N (in dbSNP:1056602).
FT FT /FTID=VAR_011692.
FT FT R->Q: LOSS OF CLEAVAGE.
FT FT MUTAGEN 255 255 S->A: LOSS OF ACTIVITY.
FT FT MUTAGEN 441 441 M -> V (IN REF. 3).
FT FT CONFLICT 160 160 I -> L (IN REF. 1).
FT FT CONFLICT 242 242 E -> Q (IN REF. 1).
FT FT CONFLICT 329 329 RAD -> KAN (IN REF. 1).
FT FT CONFLICT 489 491
FT FT SEQUENCE 492 AA; 53891 MW; CAB44FD174A9076B CRC64;
SQ Query Match 99.9%; Score 2714; DB 1; Length 492;
Best Local Similarity 99.8%; Pred. No. 2,6e-193;
Matches 491; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNCGSPPAIGFYENHGQENPYPAQPTVPTVYEHPAQYPSFVQYAPRVLTQA 60
DB 1 MALNCGSPPAIGFYENHGQENPYPAQPTVPTVYEHPAQYPSFVQYAPRVLTQA 60

```

RESULT 2

```

TMS2_MOUSE
ID TMS2_MOUSE STANDARD; PRT; 490 AA.
AC Q9J1Q8; Q9J1K4; Q9QY82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic
DE transmembrane protein X).
GN TMPRSS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=21104370; PubMed=11169526;
RX STRAIN=BALB/c;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikko P.T.;
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=BALB/c;
RX MEDLINE=20148617; PubMed=10683448;
RA Jacquinet E.J., Rao N.V., Rao G.V., Hoidal J.R.;
RT "Cloning, genomic organization, chromosomal assignment and expression
RT of a novel mosaic serine proteinase: epitheliasin.";
RL FEBS Lett. 468:93-100 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC MEDLINE=129; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;

```

FT	ACT_SITE	294	294	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	343	343	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	439	439	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	SITE	253	254	CLEAVAGE (POTENTIAL).
FT	DISULFID	76	125	BY SIMILARITY.
FT	DISULFID	119	138	BY SIMILARITY.
FT	DISULFID	132	147	BY SIMILARITY.
FT	DISULFID	171	230	BY SIMILARITY.
FT	DISULFID	184	240	BY SIMILARITY.
FT	DISULFID	243	363	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	279	295	BY SIMILARITY.
FT	DISULFID	408	424	BY SIMILARITY.
FT	DISULFID	435	463	BY SIMILARITY.
FT	CARBOHYD	111	111	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	212	212	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	474	474	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	122	122	S -> L (IN REF. 3).
FT	CONFLICT	178	178	S -> N (IN REF. 3).
FT	CONFLICT	320	320	Y -> H (IN REF. 1).
FT	CONFLICT	474	474	N -> D (IN REF. 1).
FT	SEQUENCE	490 AA;	53479 MW;	07D2B03EA4D8A1A9 CRC64;

Query Match	77.98;	Score	2117;	DB 1;	Length	490;			
Best Local Similarity	78.4%;	Pred. No.	3.4e-149;						
Matches	385;	Conservative	41;	Mismatches	63;	Indels	2;	Gaps	2;

QY	1	MALNSGPPAIGPYENHGYQENPYPAQPTVVFTVYEVHPAQYQYPSYPVQAPRVLTQA	60
DB	1	MALNSGPPPGPGCYENHGYQSEHICPPRPVPAPNGYNLYPAQYQYPSYPVQAPRITTOA	60
QY	61	SNPVVCTOPKSPSGTCTSKTKALCITLTGLTFLYGAALAGLWKPNGSKCSNSGIRC	120
DB	61	STSVIHTPKS-SGAPCTSKSKSLCLALAGVLTGAAVAVLLWRFWDSNCSISEMEC	119
QY	121	DSSTCINPNWCODGVSHCPGGSEENRCVRLYGNFIQVYSSQRKSWHPVQDDNNENY	180
DB	120	GSSGTCTSSSLWCDGVGACHPGSEENRCVRLYQSGFIQVYSSQRKAMYPVQDDWSEY	179
QY	181	GRAACRDWGYKNFYSSQGVDDSGSTSPMKLNTSAGNVDYKYLHSDACSKAVLSLR	240
DB	180	GRACKDGYKNFYSSQGVDDSGSTSPMKLNTSAGNVDYKYLHSDACSKAVLSLR	239
QY	241	CIACGVNLNSRSRIRIVGSGSALPGAWPMQVSLHVNQNVHVCQGSITPEWIVTAACHVEK	300
DB	240	CIECGVR-SVKROSRIYGGINAGSGDWPQVSLHVQGVHVCQGSITPEWIVTAACHVEE	298
QY	301	PLANNPWTATFAGILROSEMFYAGYQVEKVIISHPNYDSKTKNDIALMKLOKPLTFNDL	360
DB	299	PLSGPRWTATFAGILROSLNFYSRQVEKVIISHPNYDSKTKNDIALMKLOTLPLAFNDL	358
QY	361	VKPVCLNPGMMLQPEOLCWSWGATGEEKTKTSEVLNAKVLIIETORCSNRYVTDMLI	420
DB	359	VKPYCLENPGMMLDQECWISWGATYFKGKTSDVLNAAMVPLTEPKSCNSKIYNLLI	418
QY	421	TPAMICAGFLQGNVDSQCQSDSGGPLYTSKNNTIWLIGDTSWGSQGCAYRPGVGNVWF	480
DB	419	TPAMICAGFLQGSVDSQCQSDSGGPLYTLKNGIWLIGDTSWGSQGCAYRPGVGNVWF	478
QY	481	TDWITYROMRAD	491
DB	479	TDWITYQOMRAN	489

RESULTS	3
TMS3	MOUSE
ID	TMS3
MOUSE	STANDARD;
Q8K1T0	Q8VDE0;
AC	10-OCT-2003 (Rel. 42, Created)
DD	10-OCT-2003 (Rel. 42, Last sequence update)
DT	10-OCT-2003 (Rel. 42, Last sequence update)
DE	15-MAR-2004 (Rel. 43, Last annotation update)
GN	Transmembrane protease, serine 3 (BC 3.4.21.-).
OS	TMPSR33.
Mus musculus	(Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND FUNCTION IN ENAC
 RP CLEAVAGE.
 RX MEDLINE=22281255; PubMed=12393794;
 RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
 RA Doughtery L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
 RA Bucher K., Raymond A., Hummler E., Marzella P.L., Kudoh J.,
 RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier S.C.;
 RT The transmembrane serine protease (TMPRSS3) mutated in deafness
 RT DNFB8/10 activates the epithelial sodium channel (ENaC) in vitro.;
 RL Hum. Mol. Genet. 11:2829-2836(2002).
 CC -!- FUNCTION: Probable protease. Seems to be capable of activating
 CC ENaC.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
 CC reticulum.
 CC -!- TISSUE SPECIFICITY: Expressed in the spiral ganglion, the cells
 CC supporting the organ of Corti and the stria vascularis.
 CC -!- PTM: Undergoes autolytic activation.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
 CC -!- SIMILARITY: Contains 1 SRCR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AJ429216; CAD22137.1; --
 DR EMBL; AJ300738; CAC83350.1; --
 DR KSSP; P00761; IAN1.
 DR MGD; MGI:2155445; Tmprs3.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001130; Ssr_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLa; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; PS01209; LDLRA_1; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS00568; LDLRA_2; 1.
 DR PROSITE; PS0287; SRCR_2; 1.
 DR PROSITE; PS0240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 DR Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
 KW Endoplasmic reticulum.
 FT DOMAIN 1 48
 FT TRANSMEM 49 69
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT LDL-RECEPTOR CLASS A.
 FT SRCR.
 FT SRPINE PROTEASE.
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 257 257
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 304 304
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 215 217
 FT CLEAVAGE (POTENTIAL).
 FT BY SIMILARITY.
 FT DISULFID 73 85
 FT BY SIMILARITY.
 FT DISULFID 79 98
 FT BY SIMILARITY.
 FT DISULFID 92 107
 FT BY SIMILARITY.
 FT DISULFID 129 194
 FT BY SIMILARITY.
 FT DISULFID 142 204
 FT BY SIMILARITY.
 FT DISULFID 207 324
 FT BY SIMILARITY.

FT DISULFID 242 258 BY SIMILARITY.
 FT DISULFID 338 406 BY SIMILARITY.
 FT DISULFID 369 385 BY SIMILARITY.
 FT DISULFID 396 424 BY SIMILARITY.
 FT CARBOHYD 221 221 N-LINKED (GLCNAC... (POTENTIAL).
 FT CONFLICT 117 117 L -> H (IN REF. 1; CAC83350).
 FT CONFLICT 246 246 V -> I (IN REF. 1; CAC83350).
 SQ SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF6E1BF6 CRC64;
 Query Match 32.4%; Score 880; DB 1; Length 453;
 Best Local Similarity 42.9%; Pred. No. 8.5e-58;
 Matches 187; Conservative 64; Mismatches 159; Indels 26; Gaps 12;
 QY 69 PKPSGTVCTSTKTKALCI-----TLTLGLTFLVGAALAGLLWFKMGSKNSGIEDSSG 124
 DB 28 PVAPDGAQAQILSLPLKFFPIIVIGIILALALALGGLHFE---DCSGK-YRCHSSF 83
 QY 125 TCINPSNWCDSVSHCPGGEDENRVLYGNFVLOVYSSQKSHHPVCCDDNNENYGRAA 184
 DB 84 KCIELTARCDGSDCKNAEDYRCVRSORALQVFTA--AAWRTMCSDDMKSHYAKIA 441
 QY 185 CRDMGYKNFYSSQGI---VDPSSGTSFMKLNTSAGNVDIYKLYHS----DACSSKAV 236
 DB 142 CAQLGFP-SYVSDHLRVDALAEQFGDFVSIINELLSLSD-DKVTALHHSVYMRGCTSGHV 199
 QY 237 VSLRIACGVNLSSRSRIRVGGESALPGAPWQVSHVQNVHVCVGSIIITPEWIVTAAH 296
 DB 200 VILKKSACGRTGYS--PRIVGNNMSLTQWPPQVSLQPGYHLCGGSVITPLWIVTAAH 257
 QY 297 CVEKPLNNPWHMTAFAGILRQSFMYGA-GYQVEKVISHPNYDSKTKNNDIAMKLQKPL 355
 DB 258 CV-YDLVHPKSWTVQVGLV--SLMDSVPVSHLVKEIYHSHYKPKPLGNDIALMKLSEPL 314
 QY 356 TFNDLVKPVCLPNPQMLQPEQLCSGWCATBEKGTSEVLNAAKVLLETQRCNSRYV 415
 DB 315 TFDTIQICLIPNSEENFPDGKLCWTSWGATDGGDAPVLNHAAPVLSINKIENRDV 374
 QY 416 YDNLTITPAMICAGFLQNVDSQCGDGLVTSKNNIWLIGTSMGSGCAKAYRPGVYG 475
 DB 375 YGGIISPSMLCAGYKGGVDSQCGDGLVTCQERBLWKLVGATSGFGICGAENVKPGVYT 434
 QY 476 NVMVFTDVIYQMRAD 491
 DB 435 RITSFLOWIHEQLERD 450
 RESULT 4
 TMS3 HUMAN
 ID TMS3 HUMAN STANDARD; PRT; 454 AA.
 AC P57727.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Transmembrane protease, serine 3 (EC 3.4.21.-) (Serine protease
 TADG-12) (Tumor associated differentially-expressed gene-12 protein).
 GN TMPRSS3 OR TADG12 OR ECHOS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND T).
 RC TISSUE=Ovarian carcinoma;
 RX MEDLINE=20521358; PubMed=11068177;
 RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
 RA Wang Y., Farnley T.H., O'Brien T.J.;
 RT Ovarian tumor cells express a novel multi-domain cell surface serine
 RT protease.;
 RL Biochim. Biophys. Acta 1502:337-350(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND D), AND VARIANT IL2-53.
 RX MEDLINE=20578749; PubMed=1137999;
 RA Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chraet R.,

-/- SIMILARITY: Contains 1 SROR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>) or send an email to license@isb-sib.ch.

EMBL; AF201380; AAG37012.1; -
DR EMBL; AB038157; BAB20077.1; -
CC EMBL; AB038158; BAB20078.1; -
CC EMBL; AB038159; BAB20079.1; -
CC EMBL; AB038160; BAB20080.1; -
CC HSSP; P00763; LDPO. -
DR MEROPS; S01.079; -
DR Genew; HGNC:11877; TMRPS3.
DR MIM; 605511; -
DR MIM; 601072; -
DR MIM; 605316; -
DR GO; GO:0016021; C:integral to membrane; NAS.
GO; GO:0004252; F:serine-type endopeptidase activity; NAS.
GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
PFam; PF00057; ldl_recept_a; 1.
PFam; PF00089; tryptsin_1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SMART0192; LDLra; 1.
DR SMART; SMART0202; SR; 1.
DR SMART; SMART0020; TRYP_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00068; LDLRA_2; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00133; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW Endoplasmic reticulum; Deafness; Alternative splicing;
KW Disease mutation; Polymorphism.
DOMAIN 1 48
TRANSMEM 49 69
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL)
EXTRACELLULAR (POTENTIAL).
LDL-RECEPTOR CLASS A.
SRCR.
SERINE PROTEASE.
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CLEFTAGE (POTENTIAL).
BY SIMILARITY.
85
BY SIMILARITY.
79 98
BY SIMILARITY.
92 107
BY SIMILARITY.
129 194
BY SIMILARITY.
142 204
BY SIMILARITY.
207 324
BY SIMILARITY.
242 258
BY SIMILARITY.
338 407
BY SIMILARITY.
370 386
BY SIMILARITY.
397 425
BY SIMILARITY.
N-LINKED (GLCNAC...) (POTENTIAL).
MISSING (in isoform B).
/FTID=VSP_005391.
VAESPPLIC 318 454
EMIQPVCLNSBNFDPGKVQWTSWGKATDGDASPVLN
HAIQPLSNKNENHDYVGIIISPMLCAGYLITGGDVSCQG
DSGGPIVCVLKATSGFGICAEVNKEGVTVRTVSF
LDWIHOMERDUKT > GTSGSLGSAALPFODLIOLLI

FT	TRANSMEM	31	51	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)	RN
FT				(POTENTIAL)	RP
FT	DOMAIN	52	435	EXTRACELLULAR (POTENTIAL)	RC
FT	DOMAIN	59	101	LDL-RECEPTOR CLASS A.	RX
FT	DOMAIN	102	202	SRCL.	RA
FT	DOMAIN	203	435	SERINE PROTEASE.	RA
FT	ACT_SITE	243	288	CHARGE RELAY SYSTEM (BY SIMILARITY).	RT
FT	ACT_SITE	268	243	CHARGE RELAY SYSTEM (BY SIMILARITY).	RT
FT	ACT_SITE	385	385	CHARGE RELAY SYSTEM (BY SIMILARITY).	RL
FT	ACT_SITE	202	202	CLEAVAGE (POTENTIAL).	RN
FT	SITE	62	81	BY SIMILARITY.	RP
FT	DISULFID	75	90	BY SIMILARITY.	RA
FT	DISULFID	125	181	BY SIMILARITY.	RT
FT	DISULFID	138	191	BY SIMILARITY.	RT
FT	DISULFID	134	308	BY SIMILARITY.	RT
FT	DISULFID	228	244	BY SIMILARITY.	RL
FT	DISULFID	334	370	BY SIMILARITY.	RP
FT	DISULFID	381	408	BY SIMILARITY.	RA
FT	CARBOHYD	128	128	N-LINKED (GLCNAC. .) (POTENTIAL).	RT
FT	CARBOHYD	176	176	N-LINKED (GLCNAC. .) (POTENTIAL).	RL
SQ	SEQUENCE	435 AA;	47495 NW;	DC52E45A43E01369 CRC64;	RN
Query Match					25.0%; Score 680; DB 1; Length 435;
Best Local Similarity					36.8%; Pred. No. 4.9e-43;
Matches 163; Conservative					61; Mismatches 173; Indels 46; Gaps 13;
QY		68	OPKPSGIVCTSKTKA-----LCITLTLTGTLVGAALAGLWFKMGKSCNSGIECDSS	123	RA
DB		19	KPRRQETP-----KKVGPIIAVLLSLIALIVALLIKVILDKY-----FICGSP	65	RA
QY		124	GTCINPSNWDGVSHCPGGEENRC-----VRLYGNFTLOVYSQKSHWPCV	172	RA
DB		66	LTFIQRLCDGHLDCASGEDEHCVKDFPEKPGVAVRLSKDRSTQLVDLATGTWASVC	125	RA
QY		173	ODMWNENYGRACRDMGKYN--FYSSQGIIVDSDGSTSFMKLNTSAGNVDIYKLYHSDAC	231	RA
DB		126	FDNFTALAKTACRQGVDSQAPFAVEIRPDN---LPVAQVTGNSQELQVQNGSRSC	181	RA
QY		232	SKAVVSLRLCIAGVNLSSROSRIVGESALPGAWPQVSLHVQNVHVCSSIIITPMI	291	RA
DB		182	LSGSLVSLRCLDCGKSL---KTPRVGVGVEAPVDSMPWQVSIQYNKQVCGSILDPHWI	238	RA
QY		292	VTAACHCVRPLN-NPWHWTAFAIGLRQSPFMFYGAGVQVEKV--ISHPNYDSTKTNNDIALM	349	RA
DB		239	LTAACHFRKLVDSVSKVAGNVLGNS-----PSLPVAKIFIAEPN-PLYPKEDIALV	292	RA
QY		350	KLQKPLTFNDLVKPCPLNPGWMLQPEQLCWISGWATEEK-GKTSEVLNAAKVLIIQTQ	408	RA
DB		293	KLQMLPTFGSVRPICLPFSDEVLPATPVVWIGWGFTTEENGKMGKMDLQASVQVIDST	352	RA
QY		409	RNSRVYVNLITPAMICAGFLQGNVDSQGDSPGLVTSKNNIWLIGDTSWGSGCAKA	468	RA
DB		353	RCNAEDAYEGEVTAEMLCAGTPQGGKOTCGDGGGLPMTHSDK-WQVVGIVSWGHGCGGP	411	RA
QY		469	YRPGVYGNVWFTDIYRQWRAD	491	RA
DB		412	STPGVYTKVTAYLNIWYVRKSE	434	RA
RESULT 6					
TMS4 HUMAN					
ID	TMS4 HUMAN	STANDARD;	PRT;	437 AA.	
AC	C9NZ54; C9NZ55;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DE	Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine				
DE	protease 2) (MT-SP2).				
GN	TPRPS4 OR TMRPS3.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				

DR Pfam; PF00057; ldl_recept a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDla; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00202; TRyp_Spc; 1.
 DR PROSITE; PS01209; LDRA_1; FALSE NEG.
 DR PROSITE; PS00668; LDRA_2; FALSE NEG.
 DR PROSITE; PS00420; SRCR_1; FALSE NEG.
 DR PROSITE; PS00287; SRCR_2; 1.
 DR PROSITE; PS0240; TRYPsin DOM; 1.
 DR PROSITE; PS0134; TRYPsin HIS; 1.
 DR PROSITE; PS0135; TRYPsin SER; 1.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
 FT DOMAIN 1 32 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 33 53 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 54 437 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 61 93 LDL-RECEPTOR CLASS A.
 FT DOMAIN 94 204 SRCR.
 FT DOMAIN 205 437 SERINE PROTEASE.
 FT ACT_SITE 245 245 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 204 205 CLEAVAGE (POTENTIAL).
 FT DISULFID 64 83 BY SIMILARITY.
 FT DISULFID 77 92 BY SIMILARITY.
 FT DISULFID 127 183 BY SIMILARITY.
 FT DISULFID 140 193 BY SIMILARITY.
 FT DISULFID 196 310 BY SIMILARITY.
 FT DISULFID 230 246 BY SIMILARITY.
 FT DISULFID 356 372 BY SIMILARITY.
 FT DISULFID 383 410 BY SIMILARITY.
 FT CARBOHYD 130 130 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 178 178 MLDPSQDPSLNSLDVLPKRPDMETFRK -> MSNPCA
 FT CONFLICT 1 31 NPYSPWRSES (IN REF. 2).
 SQ SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;

Query Match 24.9%; Score 676.5; DB 1; Length 437;
 Best Local Similarity 39.1%; Pred. No. 96-43;
 Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 133 CDQVSHCPGGEENRCVRLV--GP-----NFILOYSSQRKSHHPVCQDDWNNY 181
 DB 77 CDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQLVLSATGNWFSACFDNFTALA 136

QY 182 RAACRDMGY--KNFYSSQ-----GIVDPSGTSFMKLNTSAGNVDIYKLYHSDACS 232
 DB 137 ETACROMGYSSKPTFAVEIGPDQLDVVEITENSQELRMNNSG-----PCL 184

QY 233 SKAVVSLRCIACGVNLNSRQRIYGGESALPGAPWQVSLHVQNVHVCSSIIITPEWIV 292
 DB 185 SGLSVLHCLACGKSL---KTPRVVGGEEASVDSWPMQVSLQYDKQKQVCGSILDPHWVL 241

QY 293 TAAHCKVEKFLNPFWHTAFAGILR--QSFMYGAGYQVEKVI---SHFNYSKTKNNIDIAL 348
 DB 242 TAAHCKPKH--TDVFNKVRAGSDKLGSP-----PSLAVAKIIIEFNPMY---PKNDIAL 293

QY 349 MKLQKPLTENDLVKPCLENPCGMLOPQLCHISGATEEK-GKTSEVLNAAKVLIET 407
 DB 294 MKLQPLTETSGTRVPCLEPFFDEELTPATPMIIGWFTKQNGKQMSDILLOQASVQVIDS 353

QY 408 QRCNSRYVDNLITFAMICAGFLOGNVDSQDGGPLVTSKNNIWMILGDTSGSCCAK 467
 DB 354 TRCNADDAVQGEVTEKQMCAGIPFEGGVDTCCQDGGPLMYQSDQ--WHVVGIVSGNGYCGCG 412

QY 468 AVRFPGYGVNMYFTWIPQWAD 491
 DB 413 PSTPGVTKVSAYLNNIYNWRAE 436

RESULT 7

TM35 MOUSE
 ID TM35 MOUSE STANDARD; PRT; 455 AA.
 AC Q9ER04; Q9ER02; Q9ER03;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
 GN TMPSRS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1] _SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RP TISSUE=Brain;
 RC TISSUE=Brain;
 RA Mitsui S., Yamaguchi N.;
 RT "cDNA cloning of mouse spinesin";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Brain;
 RA Mitsui S., Yamaguchi N.;
 RT "Molecular cloning of mouse type 4 spinesin";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC 1- SUBCELLULAR LOCATION: Type II membrane protein (potential).
 CC 1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=4;
 CC IsoId=Q9ER04-1; Sequence=Displayed;
 CC Name=1;
 CC IsoId=Q9ER04-2; Sequence=VSP_005397, VSP_005398;
 CC Name=2;
 CC IsoId=Q9ER04-3; Sequence=VSP_005395;
 CC Name=3;
 CC IsoId=Q9ER04-4; Sequence=VSP_005396;
 CC 1- SIMILARITY: Belongs to peptidase family S1.
 CC 1- SIMILARITY: Contains 1 SRCR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AB016229; BAB20276.1; -
 DR EMBL; AB016230; BAB20277.1; -
 DR EMBL; AB016423; BAB20278.1; -
 DR EMBL; AB041037; BAB40328.1; -
 DR HSSP; PC0763; IDPO.
 DR MEROPS; S01.313; -
 DR MGD; MGI:1933407; TmpRSS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRyp_Spc; 1.
 DR PROSITE; PS0240; TRYPsin DOM; 1.
 DR PROSITE; PS00134; TRYPsin HIS; 1.
 DR PROSITE; PS00135; TRYPsin SER; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0287; SRCR_2; 1.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein; Alternative splicing.
 FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 71 455 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 112 207 SRCR.
 FT DOMAIN 218 455 SERINE PROTEASE.

DR P1am; PFO01390; SEA; 1.
 DR P1am; PFO0530; SRCR; 1.
 DR P1am; PFO0089; CYP51; 1.
 DR PRINTS; PFO0722; CHYMOTRYPSIN.
 DR PRINTS; PFO0261; LDLRECEPTOR.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00192; LDLA; 2.
 DR SMART; SM00137; MAM; 1.
 DR SMART; SM00200; SEA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00202; TYD; SPC; 1.
 DR PROSITE; PFO1180; CUB; 2.
 DR PROSITE; PFO1209; LDLRA_1; 2.
 DR PROSITE; PFO0688; LDLRA_2; 2.
 DR PROSITE; PFO0740; MAM; 1; 1.
 DR PROSITE; PFO0680; MAM; 2; 1.
 DR PROSITE; PFO0420; SEA; 1.
 DR PROSITE; PFO0420; SRCR; 1; FALSE_NEG.
 DR PROSITE; PFO0287; SRCR; 2; 1.
 DR PROSITE; PFO0240; TRYPSIN_DOM; 1.
 DR PROSITE; PFO0334; TRYPSIN_HIS; 1.
 DR PROSITE; PFO0135; TRYPSIN_SER; 1.
 KW Signal-anchor; Glycoprotein; Myristate; Hydrolase; Serine protease;
 KW Zymogen; Transmembrane; Repeat; Alternative splicing; 3D-structure;
 KW Lipoprotein.

1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
 1 18 CYTOPLASMIC (POTENTIAL).
 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 48 1035 EXTRACELLULAR (POTENTIAL).
 54 169 SEA.
 197 238 LDL-RECEPTOR CLASS A 1.
 240 350 CUB 1.
 358 520 MAM.
 540 650 CUB 2.
 657 695 LDL-RECEPTOR CLASS A 2.
 694 787 SRCR.
 801 1035 SERINE PROTEASE.
 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).
 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).
 987 987 CHARGE RELAY SYSTEM (BY SIMILARITY).
 2 2 N-myristoyl glycine (Potential).
 199 212 BY SIMILARITY.
 206 225 BY SIMILARITY.
 219 236 BY SIMILARITY.
 659 671 BY SIMILARITY.
 665 684 BY SIMILARITY.
 678 693 BY SIMILARITY.
 788 912 INTERCHAIN (BY SIMILARITY).
 826 842 BY SIMILARITY.
 926 933 BY SIMILARITY.
 957 972 BY SIMILARITY.
 983 1011 BY SIMILARITY.
 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).
 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
 698 698 N-LINKED (GLCNAC. . .) (POTENTIAL).
 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
 762 762 N-LINKED (GLCNAC. . .) (POTENTIAL).
 864 864 N-LINKED (GLCNAC. . .) (POTENTIAL).
 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 166 192 Missing (in isoform Short).
 FT CONFLICT 808 808 /FTID=VSP_005386.
 SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;
 Query Match 24.3%; Score 660; DB 1; Length 1035;
 Best Local Similarity 33.9%; Pred. No. 4.1e-41;
 Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
 QY 65 VCTQPKSPSGTVCTSKTKKALCTLT-----GFLVCAALAAAGLLWFKMGSKCSNSG 117
 DB 611 VTGPGPVNDVFTSTNRTVLFITDNLAKQGFANFTTGYGLG-----IPECKEDN 663
 QY 118 IECDSSTGCTINFSNWCDSVSHCPGDEENRCVLYG-----PNFLLQVSSQSKSWH 169
 DB 664 FQC-KDGEICPLVNLCDGFPCHKXGDSDEAHCVELFNGTTDSSGLVQFRIQ-----SIWH 716
 QY 170 PVCQDDNNENYGRAACRDMGYKNFYSSQGVIVDSSTSPMKLNTSAGNVDIYKLYHSD 229
 DB 717 VACAENWTQISDDVQLLGLTG--NSVPTFTSTGGPYVNLNTAPNSLI---LTPSQ 771
 QY 230 ACSSKAVVSLRC--IACGVNLSRQS-RIVGSESALPGAWPQVSLHVQNVHVCSSII 286
 DB 772 QCLEDSLILLQCNVYKSCGKKLVQEVSPKIVGSDSREGAWPWVALYFDDQVCGASLV 831
 QY 287 TPEWIVTAACHVEXPLANNPHWTAFAGILAQSPMFYAGYQVE-----KVISHPNYDSKT 341
 DB 832 SRDLVLSAAHCYVGRNWEPSKXAVUGLHNASNL---TSPQIETRLIDQIVINPHYNKR 888
 QY 342 KNDIALMKLQKPLTFNDLVKPLNPGMMLQFEOLCWISGNGATEEKGKTSVLNAAK 401
 DB 889 KNDIAMHLEKVNVTYDIQICLPEENQVFPFGKICSIAGWALYIQGSTADVLQEAQ 948
 QY 402 VLLIETORCSRYVYVNLITPAMICAGFLOGNVDSQSGGGLVTSKNNKWLIGDTSW 461
 DB 949 VFLLSNEKQQQQMPEYN-ITENMVCAGYEAGGVDSQSGGGLVTSKNNKWLIGDTSW 1007
 QY 462 GSGCAKAYRPGVYGNVMVFTDWI 484
 DB 1008 GYQCALPNRPGVYARVPRFTWI 1030
 RESULT 9
 TM55 HUMAN
 ID TM55 HUMAN STANDARD; PRT; 457 AA.
 AC Q9H3S3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
 GN TMPSRS5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX PubMed=11741866;
 RT Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
 RT "Spinesin/TMPRSS5, a novel transmembrane serine protease, cloned from
 human spinal cord.";
 RL J. Biol. Chem. 277:6806-6812(2002).
 CC -|- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -|- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
 CC neurons, in their axons, and at the synapses of motoneurons in the
 CC spinal cord.
 CC -|- SIMILARITY: Belongs to peptidase family S1.
 CC -|- SIMILARITY: Contains 1 SRCR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/> or send an email to license@ebi.ac.uk).

EMBL; AB028140; BAB20375.1; --
 HSPF; P00763; IDEO; TMRSS5.
 Genew; HGNC:14908; TMRSS5.
 MIM; 606751; --
 MEROPS; S01.313; --
 InterPro; IPR009003; Cys Ser trypsin.
 InterPro; IPR001254; Peptidase S1.
 InterPro; IPR001314; Peptidase_S1A.
 InterPro; IPR001190; S1cr_receptor.
 Pfam; PF00089; trypsin; 1.
 PRINTS; PR00722; CHYMOTRYPSIN.
 SMART; SM0020; Tryp_SPC; 1.
 PROSITE; PS0240; TRYPSIN_DOM; 1.
 PROSITE; PS00134; TRYPSIN_HIS; 1.
 PROSITE; PS00135; TRYPSIN_SER; 1.
 PROSITE; PS00420; S1cr_1; FALSE_NEG.
 PROSITE; PS0287; S1cr_2; FALSE_NEG.
 Hydrolase; Serine protease; Transmembrane; Signal-anchor;
 Glycoprotein.
 DOMAIN 1 49 CRYPTOPLASMIC (POTENTIAL).
 TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 DOMAIN 71 457 EXTRACELLULAR (POTENTIAL).
 DOMAIN 112 207 S1cr.
 DOMAIN 218 457 SERINE PROTEASE.
 ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
 ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
 ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SITE 217 218 CLEAVAGE (POTENTIAL).
 DISULFID 135 196 BY SIMILARITY.
 DISULFID 148 206 BY SIMILARITY.
 DISULFID 209 328 BY SIMILARITY.
 DISULFID 243 259 BY SIMILARITY.
 DISULFID 374 390 BY SIMILARITY.
 DISULFID 401 429 BY SIMILARITY.
 CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SEQUENCE 457 AA; 49574 MW; 64406AB495A2651 CRC64;

Query Match 24.0%; Score 651.5; DB 1; Length 457;
 Best Local Similarity 32.1%; Pred. No. 6.7e-41;
 Matches 161; Conservative 63; Mismatches 194; Indels 83; Gaps 16;
 QY 8 PPAIGPYENHGYQENPVPAQPTVVPTVYVHPA-QYPSFVPQYAPRVLTAQSNPVVC 66
 Db 9 PPMQAQYAE-----EGCPG-----IPRAEPDQHP-----ISQA----- 39
 QY 67 TOPKSPSTGVTSTKTKALCTLTGLTFLVGAALAGLLWFMGSKCN--SGTECDSSG 124
 Db 40 -----VWRSMRRCVAVLGAIG-LLAGAGVSGWLLVLYLCPAASQPSITGLQDEEI 89
 QY 125 TCINFSNCDGSHCPGCGEDENRCV-----RLVGNPFILOVYSSQKSWHPVVCDD 175
 Db 90 TL-----SCSEASAEALLPALPKTVFINSDELLEAQVEDQPRLLVCHEG 138
 QY 176 WNEYGRAACRDMGY-KNFFYSSQGVDSSTSPKWLNTAGNVDIYKKI--YHSDA-- 230
 Db 139 WSPALGLQICWBLGHLRLTHKGNVLT-----IKLNSQFAQLSPRLGFLAEAWQ 191
 QY 231 ----CSSKAVSLRTGACVNLNSRQSRIVGSGALPCAMPVQVSLHVQNVHVCQGSII 286
 Db 192 PRNCTSGQVSLRSECAGR----PLAGRIVGQSVAPGRWPQASVALGFRHCTCGGVL 248
 QY 287 TPWEIVTAAHCVEK-PLNPNPHTWATFAGILRQSFMYGAGYQVEKVISHPNYSKTKND 345

Db 249 AFRWVTAACHMSFRLARLSSWRVHAGLVSHSAVRPHGALVERIIPHPLYSAQNHDYD 308
 QY 346 IALMKLQKELTNDLVKVPCLNPNPGMLQPEOLCWISGNGATEEKQK-TSEVLNAKAVLL 404
 Db 309 VALLRLQTLNFSNTVGAVCLPAKQHPFKSGRCVSGWGHGTHPSHTYSSDMLQDTVPPL 368
 QY 405 IETORCNSRYVYDNLITPMICAGELQGNVDSQCGSPVLTSTKNINWLLGDTISWGS 464
 Db 369 FSTOLCNSCVYSGALTTPMLCAGILDRADACQDQSGGLVCPDGDTRLVGVVSGRA 428
 QY 465 CAXAYRPGVYGNVMVFTDWIY 485
 Db 429 CAEPNHPGVYAKVAEFLDWIH 449

RESULT 10

ENTK_HUMAN STANDARD; PRT; 1019 AA.
 ID ENTK_HUMAN
 AC P98073;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Enteroproteptidase precursor (EC 3.4.21.9) (Enterokinase).
 GN PRSS7 OR ENTK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ductenium;
 RA MEDLINE=95234679; PubMed=7718557;
 RX Kitamoto Y., Veille R.A., Denis-Keller H., Sadler J.E.;
 RT "cDNA sequence and chromosomal localization of human enterokinase,
 RL the proteolytic activator of trypsinogen.";
 RN Biochemistry 34:4562-4568(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND DISEASE.
 RX MEDLINE=21606074; PubMed=11719902;
 RA Holzinger A., Maier E.M., Buck K., Mayerhofer P.U., Kappler M.,
 RX Wauthrich J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roscher A.A.;
 RT "Mutations in the proenteropeptidase gene are the molecular cause of
 RL congenital enteropeptidase deficiency.";
 RN Am. J. Hum. Genet. 70:20-25(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Mense U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudo J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minooshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloescher H.,
 RA Ranser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [4]
 RP SEQUENCE OF 749-1019 FROM N.A.
 RC TISSUE=Ductenium; PubMed=8052624;
 RX MEDLINE=94329561; PubMed=8052624;
 RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
 RL protease composed of a distinctive assortment of domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 CC -I- FUNCTION: responsible for initiating activation of pancreatic
 CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
 CC A). It catalyzes the conversion of trypsinogen to trypsin which in

[illegible]

```

Query Match      23.8%; Score 647; DB 1; Length 1019;
Best Local Similarity 35.8%; Pred. No. 3.7e-40;
Matches 138; Conservative 67; Mismatches 154; Indels 26; Gaps 10;

QY 113 CSNSGIECDSCGTINPNWCDGSHCPGEGDENRCVRLYG-----PNFLLQVSSQ 164
DB 543 CKADHFC-KNGECVPLVNICDGLHCEGDSBACVRFNGFTNNGLVRFRIQ-----696
QY 165 RKSHPYQDDNENYGRACRDMGKYNFYSSQIGVDDSGSTFNMKLTSAAGNVDIYK 224
DB 697 -SIWHTACAEWTTQISNDVCQLLGLSG-NSSKPIFSTDGG-PFVKLTAP---DGLH 750
QY 225 LYHSDACSSXAVSLRC--TACGVNLNSR-QSRIVGGSALPCAMPWQVSHVONHVC 281
DB 751 LTPSQQLQSLRLQCNHKSCKGLAODITPKIVGSSNAKEGAMPWVGLYCGRLLC 810
QY 282 GGSITPEWITVAACHCEKPLNPNWHTAFAGILRQSFNF--YGAGYQVEKVISHPNYDS 339
DB 811 GASLVSSDLVSAACHVYGRNLEPSKWTALIGLEMSNLTSQTVPRLIDEIVINPHYR 870
QY 340 KTKNDIATMLKLOKPLTFNDLVKPLVNPVGMLOPEQLCWISGWGATEKTKSEVLNA 399
DB 871 RKKNDIAMHLEFKYNYTYIYQICLPEENQVFPFGRNCISAGWGVVYQGTANILQE 930
QY 400 AKVLLIETORCNSRYVDNLTITPAMICAGFLOGNVDSQDGGGPLVTSKNINWLGDT 459
DB 931 ADVPLLNSERCCQOMPEYN-ITENMICAGYEGLDSCQDGGGGLMCENNWFLAGVT 989
QY 460 SWGSGCKAVRPGVGNVWFTDWI 484
DB 990 SFGYKCALPNRPGVYARVERFTWI 1014

RESULT 11
ENTK_MOUSE
ID _ENTK_MOUSE STANDARD; PRT: 1069 AA.
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Enteropeptidase (EC 3.4.21.9) (Enterokinase).
GN PRS7 OR ENTK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Ductenum;
RX MEDLINE=98147142; PubMed=9486188;
RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
RT "Structure of murine enterokinase (enteropeptidase) and expression in
RT small intestine during development.";
RL Am. J. Physiol. 274:G342-G349(1998).
CC -!- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases (By similarity).
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
CC trypsinogen.
CC -!- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multidomain (heavy) chain linked by a disulfide bond (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 1 NAM domain.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -!- SIMILARITY: Contains 1 SRCR domain.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U73378; AAB37317.1; -.
HSP; Q07954; ICR8.
MGD; MG11197523; Prss7.
InterPro: IPR000859; CUB.
InterPro: IPR009003; Cys Ser trypsin.
InterPro: IPR002172; LDL_receptor_A.
InterPro: IPR000998; NAM domain.
InterPro: IPR001254; Peptidase_S1.
InterPro: IPR001344; Peptidase_S1A.
InterPro: IPR009020; Protease_inhib.
InterPro: IPR000082; SEA domain.
InterPro: IPR001190; Srcf_receptor.
Pfam: PF00431; CUB; 2.
Pfam: PF00057; ldl_recept_a; 2.
Pfam: PF00629; NAM; 1.
Pfam: PF01390; SEA; 1.
Pfam: PF00530; SRCR; 1.
Pfam: PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
PRINTS; PR00261; LDLRECEPTOR.
PRINTS; PR00020; NAMDOMAIN.
SMART; SM00042; CUB; 2.
SMART; SM00192; LDLA; 2.
SMART; SM00137; NAM; 1.
SMART; SM00200; SEA; 1.
SMART; SM00202; SR; 1.
PROSITE; PS01180; CUB; 2.
PROSITE; PS01209; LDLRA_1; 2.
PROSITE; PS00668; LDLRA_2; 2.
PROSITE; PS00740; NAM_1; 1.
PROSITE; PS00660; NAM_2; 1.
PROSITE; PS0024; SEA; 1.
PROSITE; PS00420; SRCR; 1.
PROSITE; PS00287; SRCR_2; 1.
PROSITE; PS00340; TRYPsin_DOM; 1.
PROSITE; PS00135; TRYPsin_SER; 1.
Signal-anchor; Glycoprotein; Hydrolase; Serine protease; Zymogen;
Transmembrane; Repeat.
CHAIN 1 829
CHAIN 830 1069
TRANSMEM 19 47
DOMAIN 48 1069
DOMAIN 52 169
DOMAIN 227 268
DOMAIN 270 379
DOMAIN 387 549
DOMAIN 569 679
DOMAIN 686 724
DOMAIN 723 816
DOMAIN 730 1069
ACT_SITE 874 874
ACT_SITE 925 925
ACT_SITE 1021 1021
DISULFID 229 242
DISULFID 236 255
DISULFID 249 266
DISULFID 688 700
DISULFID 695 713
DISULFID 707 722
NON-CATALYTIC CHAIN (HEAVY CHAIN).
CATALYTIC CHAIN (LIGHT CHAIN).
CYTOPLASMIC (POTENTIAL).
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
EXTRACELLULAR (POTENTIAL).
SEA.
LDL-RECEPTOR CLASS A 1.
CUB 1.
NAM.
CUB 2.
LDL-RECEPTOR CLASS A 2.
SRCR.
SERINE PROTEASE.
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.

```



```

FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
FT DISULFID 859 975 BY SIMILARITY.
FT DISULFID 959 1027 BY SIMILARITY.
FT DISULFID 991 1006 BY SIMILARITY.
FT DISULFID 1017 1045 BY SIMILARITY.
FT CARBOHYD 147 147 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 579 579 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 675 675 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 727 727 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 770 770 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 791 791 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 999 999 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1069 AA; 118735 MW; 562549E463743C3D CRC64;

Query Match. 23.5%; Score 638; DB 1; Length 1069;
Best Local Similarity 30.5%; Pred No. 1.8e-39;
Matches 164; Conservative 92; Mismatches 196; Indels 86; Gaps 21;

QY 18 HGQPNPYPADPTVVPT-----VVEV-HPAQYPSVPVQVAPRVLTQAS---- 61
DQ 544 NGICQSQYP-EPTLVTPPELPTDCGPFELMEPNSTFSS--PNFPDKYPNQASCIWN 600
QY 62 -----NPV-----VCTQPKSFGTCTKTKAL 85
DQ 601 LNAQRGKNTQHFQFDLENINDVVRDGGEDFSLILAVTGP-GPVKDLFSITNMTV 659
QY 86 CHTLGLTFLVG--AALAAGLLWKMFG--SKCSNGIECDSSGFCINPSNWCDSVSHCPG 141
DQ 660 IFTTNMETRRKFKANFTSGY---YLGPEPCQDDEFQC-KDGNICPLGNLCLDSYPRCD 715
QY 142 GEDENRCVRLY---GPNFLOVYSSQRKSHFVPCDDWENYGRAACRDMGYKNFYSS 197
DQ 716 GSDEASCVFLNGTSNNGLVQF--NIHSIWHIACAENWTTQISNEVCHLLGLGS--ANS 771
QY 198 QGIVDSGTSFMKLNSTAGNVDIYKLYHSDACSSKAVSLRC--IACGVNLSRSQS- 254
DQ 772 SMPISSTGGGPFVRVNAQPSGLI---LTPSLQCSQDSLILLQCNHKSCKEKKVTQKVP 828
QY 255 RIYGGESALPGAPMOVSLHVNVR---VCGSITITPEWITVTAHCKVEKPLNPNWHTA 310
DQ 829 KIYGGSDAQAQAPWVVALYHEDRSTDRLLCGASLVSSDLVLSAHCYVRNLDPTRTWA 888
QY 311 FAGILRQSPFYGAGYQ--VEKVISHPNYDSKTRNDIALMKLQKPLTFNDLVKPVCLPN 368
DQ 889 VLGLHMQSNLTSPQVVRVVDQIVINPHYDRRRKVNDIAMWHLFEKVNYYTDYIQICLPE 948
QY 369 PGWMLQPEQLCWSIGWATE-EKGKTSVLNAKVLLETORCKNSRYVDNLTIPAMICA 427
DQ 949 ENQIFPGTCTCIAGHYDKINAGSTVDVLKEADVPLISNEKCCQQLPEYN-ITESHICA 1007
QY 428 GFILQGVNDSQSGSGLPVTSKNTIWLIGTSMWCGCAKAYRPGVYGVNVMFTDWTY 485
DQ 1008 GYEEGIDSCQDSGGLMCOENNRWFLVGVTSFGVQCALPNHPGVYRVYSQFIENIWH 1065

RESULT 12
ENTK_PIG
ID_ENTK_PIG STANDARD; PRT; 1034 AA.
AC F98074;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Enteropetidase precursor (EC 3.4.21.9) (Enterokinase).
GN PRSS7 OR ENTK.

```

```

OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE=Ductal mucosa;
RX MEDLINE=94327548; PubMed=8051081;
RA Matsuhashima M., Ichinose M., Yanagi N., Kakei N., Tsukada S.,
RA Miki K., Kurokawa K., Tashiro K., Shiohara K., Shinomiya K.,
RA Uneyama H., Inoue H., Takahashi T., Takahashi K.,
RA "Structural characterization of porcine enteropetidase.";
RL J. Biol. Chem. 269:19976-19982(1994).
CC -!- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -!- SUBUNIT: Heterotrimer of a catalytic (light) chain, a multidomain
CC (heavy) chain, and a mini chain.
CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -!- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 1 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 1 MAM domain.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D30799; BAB06459.1; -.
CC HSP; P00763; IDPO.
CC
CC InterPro; IPR000859; CUB.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR002172; LDL_receptor_A.
CC InterPro; IPR000998; MAM_domain.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC InterPro; IPR000082; SEA_domain.
CC InterPro; IPR001190; Srcr_receptor.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF01390; SEA; 1.
CC Pfam; PF00530; SRCR; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PRO0722; CHYMOTRYPSIN.
CC PRINTS; PRO0261; LDLRECEPTOR.
CC PRINTS; PRO0020; MAMDOMAIN.
CC SMART; SM00042; CUB; 2.
CC SMART; SM00192; LDLA; 2.
CC SMART; SM00137; MAM; 1.
CC SMART; SM00200; SEA; 1.
CC SMART; SM00202; SR; 1.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS01180; CUB; 2.
CC PROSITE; PS01209; LDLRA_1; 2.
CC PROSITE; PS50068; LDLRA_2; 2.
CC PROSITE; PS00740; MAM_1; 1.
CC PROSITE; PS50060; MAM_2; 1.
CC PROSITE; PS50024; SEA; 1.

```


DR	PROSITE; PS00420; SRCR_1; FALSE_NEG.	QY	176	WNEYGRAACRDMGYQN-----FVSSQIVDDSGSTSPMKLNTSAGNVDIYKLYHSDA	230
DR	PROSITE; PS0287; SRCR_2; 1.	Db	722	WTQTSDDYCVQLLGLGTGNSSMPFFS-----SGGPEFKLNTAPNGSLI---LTASEQ	771
DR	PROSITE; PS0240; TRYPSIN_DOM; 1.	QY	231	CSKAVVSLRC--IAGVNLSSROS-RIVGESALPGAWPQVSLHVQNVHVCGGSIIT	287
DR	PROSITE; PS0134; TRYPSIN_HIS; 1.	Db	772	CREDSILLQCHKSKGKKQVAGVSPKIVGNDSEGAWPVVALYVYNGQLLGGASLVS	831
DR	PROSITE; PS00135; TRYPSIN_SER; 1.	QY	288	PEWITAAHCVEKPLNPNHWTAFAGILRQSPFMFYAGVQ-----VEKVIHPNYDSKTK	342
KW	Signal-anchor; Glycoprotein; Repeat; Lipoprotein.	Db	832	RDLVSAACHVTVGRNLEPSKWKAILGHEMTSLN---TSPQIVTRLIDIVINPHYNNRRK	888
FT	CHAIN 52 117	QY	343	NNDIALMKLQKPLTFNDLVKPCVCLPNPGLMLOPEQLCWISGMGATBEKGTSEVLNAKV	402
FT	CHAIN 118 799	Db	889	DSDIAMTHLEFPKVNVDYIQICLPEENQVFPFGICSIAGKVIYQSPADILQEADV	948
FT	CHAIN 800 1034	QY	403	LIETQRCNSRYVDNLITPAMICAGFLQGVDSQGGSLVTSKNNIWLIGTWSG	462
FT	DOMAIN 1 18	Db	949	PLLSNEKCCQQQMPYV-N-ITENMCAGYEGGIDSCQGDGSLMCLNRRWLLAGVTSFG	1007
FT	TRANSMEM 19 47	QY	463	SGCAKAYRPGVYGVNVVFTDWI	484
FT	DOMAIN 48 1034	Db	1008	YQCALPNRPGVYARVPKFTWI	1029
FT	DOMAIN 52 169				
FT	DOMAIN 197 238				
FT	DOMAIN 240 349				
FT	DOMAIN 357 519				
FT	DOMAIN 539 649				
FT	DOMAIN 656 694				
FT	DOMAIN 693 786				
FT	DOMAIN 800 1034				
FT	ACT_SITE 840 840				
FT	ACT_SITE 891 891				
FT	ACT_SITE 986 986				
FT	LIPID 2 2				
FT	DISULFID 199 212				
FT	DISULFID 206 225				
FT	DISULFID 219 236				
FT	DISULFID 658 670				
FT	DISULFID 665 683				
FT	DISULFID 677 692				
FT	DISULFID 787 911				
FT	DISULFID 825 841				
FT	DISULFID 925 992				
FT	DISULFID 956 971				
FT	DISULFID 992 1010				
FT	CARBOHYD 116 116				
FT	CARBOHYD 147 147				
FT	CARBOHYD 170 170				
FT	CARBOHYD 194 194				
FT	CARBOHYD 283 283				
FT	CARBOHYD 343 343				
FT	CARBOHYD 350 350				
FT	CARBOHYD 403 403				
FT	CARBOHYD 455 455				
FT	CARBOHYD 485 485				
FT	CARBOHYD 518 518				
FT	CARBOHYD 549 549				
FT	CARBOHYD 645 645				
FT	CARBOHYD 697 697				
FT	CARBOHYD 701 701				
FT	CARBOHYD 721 721				
FT	CARBOHYD 740 740				
FT	CARBOHYD 761 761				
FT	CARBOHYD 804 804				
FT	CARBOHYD 863 863				
FT	CARBOHYD 902 902				
FT	CARBOHYD 964 964				
SQ	SEQUENCE 1034 AA; 114776 MW; 0388C64CF64CC368 CRC64;				
Query Match 23.3%; Score 634; DB 1; Length 1034;					
Best Local Similarity 33.3%; Pred. No. 3.4e-39;					
Matches 147; Conservative 72; Mismatches 179; Indels 44; Gaps 13;					
QY	65 VCTQPSGTVCTSKTKALCIT---LTGL---TELVGALAGLWKGKSCNSG	117			
Db	610 VYTGPGVEDVSTNRMVTLFITNDALTKGFKANFTTGVHLG-----IPECKEDN	662			
QY	118 IECDSGTCINPNNWCDGVSHCPGGEDENRCVRLYG--PNFTLQVYSSQRKSWHPVCQDD	175			
Db	663 FQCE-NGECVLLVNLCDGFSHCKDGSDEAHCVRFLNGTANNGLVQFRIQSIWHTACAEN	721			

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; APL133845; AAD31850.1; -;
 DR EMBL; APL13348; AAP21966.1; -;
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.019; -;
 DR MIM; 605236; -;
 DR GO; GO:0005887; C: integral to plasma membrane; TAS.
 DR GO; GO:0004252; F: serine-type endopeptidase activity; TAS.
 DR GO; GO:0007345; P: embryogenesis and morphogenesis; TAS.
 DR GO; GO:0006629; P: lipid metabolism; TAS.
 DR GO; GO:0006508; P: proteolysis and peptidolysis; TAS.
 DR GO; GO:0008217; P: regulation of blood pressure; TAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR000024; Fz_domain.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF011392; Fz_2.
 DR Pfam; PF00057; ldl_recept_a; 6.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00063; FRI; 2.
 DR SMART; SM00192; LDLA; 3.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; TYD_SPC; 1.
 DR PROSITE; PS00038; FZ_2.
 DR PROSITE; PS01209; LDLA_1; 6.
 DR PROSITE; PS00068; LDLA_2; 7.
 DR PROSITE; PS0240; TRYPsin DOM; 1.
 DR PROSITE; PS0134; TRYPsin HIS; FALSE_NEG.
 DR PROSITE; PS00135; TRYPsin SER; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0287; SRCR_2; FALSE_NEG.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein; Repeat.
 FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 46 66 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT 67 1042 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 134 259 FZ_1.
 FT DOMAIN 268 304 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 305 340 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 341 377 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 378 415 LDL-RECEPTOR CLASS A 4.
 FT DOMAIN 450 573 FZ_2.
 FT DOMAIN 579 614 LDL-RECEPTOR CLASS A 5.
 FT DOMAIN 615 653 LDL-RECEPTOR CLASS A 6.
 FT DOMAIN 654 690 LDL-RECEPTOR CLASS A 7.
 FT DOMAIN 690 786 SRCR.
 FT DOMAIN 802 1042 SERINE PROTEASE.
 FT ACT_SITE 843 843 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 985 985 CHARGE RELAY SYSTEM.
 FT DISULFID 790 912 BY SIMILARITY.
 FT DISULFID 828 844 BY SIMILARITY.
 FT DISULFID 955 970 BY SIMILARITY.
 FT DISULFID 981 1010 BY SIMILARITY.
 FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 451 451 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 581 581 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 597 597 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 761 761 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MUTAGEN 985 985 S->A; LOSS OF ACTIVITY.
 FT CONFLICT 854 854 W -> R (IN REF. 2).
 FT CONFLICT 876 876 K -> R (IN REF. 2).
 SQ SEQUENCE 1042 AA; 116564 MW; 7705398EBB607AD2 CRC64;
 Query Match 22.1%; Score 601; DB 1; Length 1042;
 Best Local Similarity 35.3%; Pred. No. 9.4e-37;
 Matches 138; Conservative 62; Mismatches 169; Indels 22; Gaps 14;
 QY 111 SKCSNGIECDSSGTCINPNSWCDGVSHCPGEGEDENRCVRL---YGNFILQVYSORKS 167
 DB 653 SFCQDELEC-ANHACVSRDLWCDEADCDSSDSDWDCVTLISINVNSSFLMVHRAATE- 710
 QY 168 WHPVCQDDNENYGEACRDMGYKNFYSSQGIYVDDSGSTSPMKLNT---SAGNVVDIYKK 224
 DB 711 -HVCADGWQELLSQACKQMGELGEPSTKL-IDQKEPRWLTLSHNSLSNGTTLHEL 768
 QY 225 LYHSDACSSKAVSLRCLIA--CGVNLNRSROSRIYGGESALPGAWPQVSLHVQ-NVHVC 281
 DB 769 LVNGQSCESRSKISLLCTKQCGRRPAARMKRLGGRTSRPGRPWQCSLQSEPSGHIC 828
 QY 282 GGSIIITPEWIVTAACHVEKPLNPNWHTAFAGI--LRQSPMFYAGYQVEKVSHPNYDS 339
 DB 829 GCVLIARKWLVTVAHCFE-GRENAAVKVKVGINNLDHPSVFMQTRF-VKTIILHPIYSR 886
 QY 340 KTKNNDIAMKLOKPLTFNDLVKVPCLNPFOMMLQPEQLCWIGWGATEBKGTSEVINA 399
 DB 887 AVVDYDISIVELSDISSETGVVRPVPCLNPPEQMLEPTCYVITGNG--HMGKMPFKLQE 944
 QY 400 AKVLLIETQRCNSVYVDNLITPAMICAGELQGNVDSQCGDSGGLPVTSK-NNIWMLIGD 458
 DB 945 GEVRIISLEHQCS-YFDNKITITMTCAGYESGVDSQMGDSGGLPVCXEPGGRWTLFGL 1003
 QY 459 TSWSGGC-AKAYRFGVGNVNVFTDIYRQM 488
 DB 1004 TSWSGVCFCKVLGFGVSNVSYFVEWIKROI 1034
 RESULT 14
 CORI_MOUSE
 ID CORI_MOUSE STANDARD; PRT; 1113 AA.
 AC Q92319;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-
 DE converting enzyme) (Corin) (Low density lipoprotein receptor related
 DE protein 4).
 DE CRN OR LR54.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98429596; PubMed=9756624;
 RA Tomita Y., Kim D.-H., Magoori K., Fujino T., Yamamoto T.T.;
 RT "A novel low-density lipoprotein receptor-related protein with type II
 RT membrane protein-like structure is abundant in heart."
 RL J. Biochem. 124:784-789(1998).
 CC -!- FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY
 CC BETWEEN ARG-122 AND SER-123 (BY SIMILARITY).

RT disease.";

RL Genomics 69:225-234 (2000).

RN [3]

RP SEQUENCE FROM N.A., AND VARIANTS SER-143; THR-178; GLN-202; CYS-269;

RR VAL-311; ALA-358; ALA-381; PRO-442 AND GLN-560.

RA Rieder M.J., Arnel T.J., Carrington D.P., Ozuna M., Kuldanek S.A.,

RA Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.;

RL Submitted (DEC-2002) to the EMBL/GenBank/DBS databases.

RN [4]

RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.

RF MEDLINE-91152016; PubMed-198666;

RA McMullen B.A., Fujikawa K., Davie E.W.;

RT "Location of the disulfide bonds in human plasma prekallikrein: the

RT presence of four novel apple domains in the amino-terminal portion of

RT the molecule.";

RL Biochemistry 30:2050-2056 (1991).

RN [5]

RP CARBOHYDRATE-LINKAGE SITE ASN-453.

RX MEDLINE-22660472; PubMed-12754519;

RA Zhang H., Li X.-J., Martin D.B., Aebbersold R.;

RT Identification and quantification of N-linked glycoproteins using

RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";

RL Nat. Biotechnol. 21:660-666 (2003).

CC -1- FUNCTION: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It

CC activates, in a reciprocal reaction, factor XII after its binding

CC to a negatively charged surface. It also releases bradykinin from

CC HMW kininogen and may also play a role in the renin-angiotensin

CC system by converting prorenin into renin.

CC -1- CATALYTIC ACTIVITY: Cleaves selectively Arg-|-Xaa and Lys-|-Xaa

CC bonds, including Lys-|-Arg and Arg-|-Ser bonds in (human)

CC kininogen to release bradykinin.

CC -1- SUBUNIT: The zymogen is activated by factor XIIa, which cleaves

CC and the molecule into a light chain, which contains the active site,

CC and a heavy chain, which associates with HMW kininogen. These

CC chains are linked by one or more disulfide bonds.

CC -1- DISEASE: Defects in KLKB1 are the cause of Fletcher factor

CC deficiency (MIM:229000); a blood coagulation defect.

CC -1- SIMILARITY: Belongs to peptidase family S1. Plasma kallikrein

CC subfamily.

CC -1- SIMILARITY: Contains 4 apple domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M13143; AAA60153.1; ..

DR EMBL; AF232742; AAF79940.1; JOINED.

DR EMBL; AF232734; AAF79940.1; JOINED.

DR EMBL; AF232735; AAF79940.1; JOINED.

DR EMBL; AF232736; AAF79940.1; JOINED.

DR EMBL; AF232737; AAF79940.1; JOINED.

DR EMBL; AF232738; AAF79940.1; JOINED.

DR EMBL; AF232739; AAF79940.1; JOINED.

DR EMBL; AF232740; AAF79940.1; JOINED.

DR EMBL; AF232741; AAF79940.1; JOINED.

DR EMBL; AY190920; AAN84794.1; ..

DR PIR; A00921; KQHUP.

DR HSSP; P00763; LDFO.

DR MEROPS; S01.212; ..

DR Genew; HGNC:6371; KLKB1.

DR MIM; 229000; ..

DR GO; GO:0003807; F1plasma kallikrein activity; TAS.

DR GO; GO:0006508; P1proteolysis and peptidolysis; TAS.

DR InterPro; IPR000177; Apple.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR003014; PAN.

DR InterPro; IPR001254; Peptidase S1.

DR InterPro; IPR001314; Peptidase_S1a.

RFam; PF00024; PAN; 4.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PRO0005; APPLEDOMAIN.

DR PRINTS; PRO0722; CHYMOTRYPSIN.

DR SMART; SMO0223; APPLE; 4.

DR SMART; SMO0220; Tryp_Spc; 1.

DR PROSITE; PS00495; APPLE; 4.

DR PROSITE; PS02040; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.

KW Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;

KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;

KW Repeat; Polymorphism.

FT SIGNAL 1 19

FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.

FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.

FT DOMAIN 20 105 APPLE 1.

FT DOMAIN 110 195 APPLE 2.

FT DOMAIN 200 285 APPLE 3.

FT DOMAIN 291 376 APPLE 4.

FT DOMAIN 389 621 SERINE PROTEASE.

FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .).

FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .).

FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .).

FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .).

FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .).

FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.

FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.

FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.

FT DISULFID 21 104

FT DISULFID 47 77

FT DISULFID 51 57

FT DISULFID 111 194

FT DISULFID 137 166

FT DISULFID 141 147

FT DISULFID 201 284

FT DISULFID 227 256

FT DISULFID 231 237

FT DISULFID 292 375

FT DISULFID 318 347

FT DISULFID 322 328

FT DISULFID 340 345

FT DISULFID 383 503

FT DISULFID 419 435

FT DISULFID 517 584

FT DISULFID 548 563

FT DISULFID 574 602

FT VARIANT 143 143

FT VARIANT 178 178

FT VARIANT 202 202

FT VARIANT 208 208

FT VARIANT 269 269

FT VARIANT 311 311

FT VARIANT 358 358

FT VARIANT 381 381

FT VARIANT 442 442

FT VARIANT 560 560

FT SEQUENCE 638 AA; 71369 MW; E63F9C1053838FB4 CRC64;

N -> S (common polymorphism).

A -> T.

/FTID=VAR_013596.

/FTID=VAR_016280.

H -> Q.

/FTID=VAR_013599.

H -> P.

/FTID=VAR_013600.

S -> C.

/FTID=VAR_016281.

F -> V.

/FTID=VAR_016282.

T -> A.

/FTID=VAR_016283.

S -> A.

/FTID=VAR_016284.

Q -> P.

/FTID=VAR_016285.

R -> Q.

/FTID=VAR_016286.

Query Match 21.0%; Score 569.5; DB 1; Length 638;

Best Local Similarity 32.8%; Pred No. 1; 1e-34;

Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;

QY 55 RVLTAQSNFVVTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAGLWKFWGS--- 111

Db	218	RULT--PDAFVCR-----TICTYHFN--CLFFTFYT-----NVWKIESQRNV	255
Qy	112	---KCSNGTECDs-----SG-----TCINP-----SNWCDGVSHCPGEGDENRCVRL	151
Db	256	CLLKTBESGTFSSSTFOENTISYSLTCKRTLPECHSKIYPGVDF--GGEELN-----	308
Qy	152	YGNFILQVYSSORKSMHPVCQDDWENYGRACRDMGVKNKNFYSSCGIIVDDSGSTS----	208
Db	309	--VTFVKGV-----NVCQE-----TCTKM--IRCQFFTYSLLPEDCKEKC	347
Qy	209	FMKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIACGVN--LNSSRSQSRIVGGESALPGA	266
Db	348	FLRLSMDGSPTRI---AYGTQSSG---YSLRLCNTGDNsvCTKTSTRIVGGTNSWGE	401
Qy	267	WPQVSLHVQ---NVHVCGSIIITPEWIVTAACHVE--KPLNNPWHWTAFAGILRQSFMY	322
Db	402	WPQVSLQVKLTAQRHLCGSLIGHQWVLTAAHCFDGLPLQDVWR--IYSGILNLSDITK	459
Qy	323	GAGY-QVEKVISHENYDSKTNNDIALMKLQPLTFNDLVKPYCLNPNPQWMLQPEQLCWI	381
Db	460	DTPFSQIKETIIHQNYKVSEGNHDIALIKQAPLNYTEFOPKICLPCKGDTSTIYTNCWV	519
Qy	382	SGWGATEEKGTSEVLNAAKVLLIETQCNRSRYVDNLIITPAMICAGFLQGNVDSQGDs	441
Db	520	TGWGFSKEKEIQNILQWNIPLVTNEECQKRY-QDYKITQRMVCAGYKEGKDACKGDs	578
Qy	442	GGPLVTSKNINWLIIGTSGSGCAKAYRPGVIGNVWVFTDWIYRQMR-ADG	492
Db	579	GGPLVCKHNGMWRUVGITSWGEGCARREQPGVYTKVAEYNDWILEKTQSSDG	630

Search completed: June 1, 2004, 14:35:07
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:20:43 ; Search time 45 Seconds

(without alignments)
1051.694 Million cell updates/sec

Title: US-09-615-285B-2

Perfect score: 2717

Sequence: 1 MALNSGSPAIGPVYENHGY.....VYGNVWFDMYRQMRADG 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78:**

2: PIR1:**

3: PIR2:**

4: PIR3:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	660	24.3	1035	1 A43090	enteropeptidase (E
2	647	23.8	1019	1 A56318	enteropeptidase (E
3	634	23.3	1034	1 A53663	enteropeptidase (E
4	577.5	21.3	1113	2 J8315	low-density lipopr
5	569.5	21.0	638	1 KQ8UP	plasma kallikrein
6	565.5	20.8	417	1 S0845	hepsin (EC 3.4.21.
7	564.5	20.8	855	2 JCT731	membrane-bound arg
8	556	20.5	638	1 KQ8UP	plasma kallikrein
9	538	19.8	638	1 KQ8UP	plasma kallikrein
10	533	19.6	625	1 KPHU1	coagulation factor
11	529.5	19.5	416	1 S3377	hepsin (EC 3.4.21
12	518.5	19.1	812	1 PLMS	plasmin (EC 3.4.21
13	501	18.4	421	1 S11674	acrosin (EC 3.4.21
14	501	18.4	1524	2 T30337	polyprotein - Afri
15	490.5	18.1	415	1 A34170	acrosin (EC 3.4.21
16	489.5	18.0	436	2 JCS759	acrosin (EC 3.4.21
17	486.5	17.9	761	2 JCS759	brain-specific ser
18	483	17.8	418	2 A37344	acrosin (EC 3.4.21
19	477	17.6	790	1 PLG38	plasmin (EC 3.4.21
20	475.5	17.5	431	2 S47538	acrosin (EC 3.4.21
21	473.5	17.4	421	2 S29599	acrosin (EC 3.4.21
22	472.5	17.4	810	2 B30848	plasmin (EC 3.4.21
23	470.5	17.3	437	2 S18407	acrosin (EC 3.4.21
24	470.5	17.3	810	2 I46260	plasmin (EC 3.4.21
25	461	17.0	343	2 A57014	proctasin (EC 3.4
26	456.5	16.8	460	2 B61545	plasmin (EC 3.4.21
27	455	16.7	812	1 PLSO	plasmin (EC 3.4.21
28	451	16.6	2616	2 S47096	nudel protein prec
29	450.5	16.6	267	2 S40006	trypsin (EC 3.4.21

RESULT 1

A43090

enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine

N;Alternate names: enterokinase

C;Species: Bos primigenius taurus (cattle)

C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 28-Apr-2003

C;Accession: A43090; A48874; A61436

R;Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.

Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A;Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease comp

A;Reference number: A43090; PMID:94329561; PMID:8052624

A;Accession: A43090

A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1035 <KIT>

A;Cross-references: GB:U09859; NID:g746410; PIDN:ARB40026.1; PID:g746411

A;Experimental source: small intestine

R;Avallie, E.R.; Renentulla, A.; Racie, L.A.; DiBlasio, E.A.; Perez, C.; Grant, K.L.;

J. Biol. Chem. 269, 23311-23317, 1993

A;Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of I

A;Reference number: A48874; NUID:94043122; PMID:8226855

A;Accession: A48874

A;Molecule type: mRNA

A;Residues: 801-1035 <LAV>

A;Cross-references: GB:L19663; NID:g416131; PIDN:AAA16035.1; PID:g416132

A;Note: parts of this sequence, including the amino end of the mature protein, were conl

R;Light, A.; Janska, H.

J. Protein Chem. 10, 475-480, 1991

A;Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.

A;Reference number: A61436; NUID:92189715; PMID:1799406

A;Accession: A61436

A;Molecule type: protein

A;Residues: 801-807, 'Y', 809-827 <LIG>

C;Comment: The mechanism of association with the membrane of the intestinal brush border

membrane attachment using a signal-anchor sequence.

C;Comment: Conversion from membrane-bound to soluble forms may involve further processin

l fide linked

C;Function:

A;Description: cleaves propeptide from trypsinogen to produce active trypsin

A;Pathway: intestinal digestive hydrolase cascade

C;Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding rei

C;Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protei

F;22-38/Domain: transmembrane #status predicted <TM>

F;52-117/Product: enteropeptidase mini chain #status predicted <MCH>

F;118-800/Product: enteropeptidase heavy chain #status predicted <HCH>

F;199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F;358-520/Domain: MAM homology <MAM>

F;542-647/Domain: C1r/C1s repeat homology <C1R>

F;659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F;844-799/Domain: scavenger receptor: cysteine-rich domain homology <SR>

F;801-1035/Product: enteropeptidase light chain #status predicted <LCH>

30 450 16.6 270 2 S56160 mast cell tryptase
31 449 16.5 275 2 S40005 trypsin (EC 3.4.21
32 449 16.5 655 1 A46688 hepatocyte growth
33 448 16.5 277 2 S35340 trypsin (EC 3.4.21
34 447 16.5 276 2 A38654 mast cell proteinase
35 446.5 16.4 420 2 A55283 acrosin (EC 3.4.21
36 445.5 16.4 810 1 PLHU plasmin (EC 3.4.21
37 444.5 16.4 455 2 A61545 plasmin (EC 3.4.21
38 443.5 16.3 266 2 S54146 trypsin (EC 3.4.21
39 442.5 16.3 458 1 S00657 apoprotein(a) (EC
40 441.5 16.2 274 2 JC4171 trypsin (EC 3.4.2
41 441 16.2 786 1 A47547 serine proteinase
42 439 16.2 273 2 A47246 trypsin (EC 3.4.2
43 438 16.1 274 2 S35339 trypsin (EC 3.4.21
44 437 16.1 275 2 S40007 trypsin (EC 3.4.21
45 436 16.0 242 2 S49489 trypsin (EC 3.4.21

F:801-1030/Domain: trypsin homology <TRY>
F:115,147,170,194,233,263,284,454,486,519,550,646,698,722,741,762,864,903,965/Binding
F:788-912,826-842,928-933,957-972,983-1011/Disulfide bonds: #status predicted
F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 24.3%; Score 660; DB 1; Length 1035;
Best Local Similarity 33.9%; Pred. No. 8e-40;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

QY 65 VCTQPKSPSTVCTSKTKALCTITL-----GTFVGAALAGLWKGKSCNSG 117
Db 611 VTGPGPVNVDFSTTRMTVLFITDNMLAKQFKANFTGYGLG-----IPBEKEDN 663
QY 118 IECDSSGTCINPNSWCDGVSHCPGGEDENRCVRLYG-----PNFLOVYSSQKSWH 169
Db 664 FQC-KDGEICPLVNLCDGPFCHKDGSDAHCVRLEFNGTIDSSGLVQFRIQ-----SIWH 716
QY 170 PVQDDNNENYGRAACRDGMYKNNFYSSQGIYDDSGTSFMKLNLSAGNVDIYKLYHSD 229
Db 717 VACAENWTTQISDDVQCLGLGTG--NSSVPFTSGGPGYVNLNTPAGSLI-----LPSQ 771
QY 230 ACSKSAVVSURC--IACGVNLSRS--RIVGGSALPGAMPQVSLHVQNVHVCSSII 286
Db 772 QCLEDSLILLCQNYKSCGKLVQEVSPKIVGSDSREGAMPVVALYFDDQVCGASLV 831
QY 287 TPEWIVTAACHVEKPLNNPHWTFAPAGILRQSFMEYAGYQVE-----KVISHENYDSKT 341
Db 832 SRDLVSAACHVYGRNWSKVAVLGLHMAISL---TSPQIETRLIDQIVINSHYNKR 888
QY 342 KNDIALMKLQKLTENDLVKVPCLPNPGMLQPLQCLWISGWGATEKGTSEVLYNAK 401
Db 889 KNDIAMHLEMKVNTDYIQICLPEENQVPPGRCISAGWALYVGGSTADVLQAD 948
QY 402 VLLIETQCNRSRVYDNLITPAMICAGFLQGVNDSCQDSGGPLVTSKNNIWLIGTWS 461
Db 949 VPLLSNEKCOQQMPEYN--ITENMVCAGYEAGGVDSQCGDSGGPLMCCQNNRLLAGVTSF 1007
QY 462 GSGCAKAYRPGVYGVNVVFTDWI 484
Db 1008 GYQCALPNRPGVYARVRFTEWI 1030

RESULT 2
A56318
enteropeptidase (EC 3.4.21.9) precursor [validated] - human
N/Alternate names: enterokinase
C/Species: Homo sapiens (man)
C/Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C/Accession: A56318; B43090
R/Kitamoto, Y.; Veille, R.A.; 1995
Biochemistry 34, 4562-4568, 1995
A/Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic
A/Reference number: A56318; MUID:95234679; PMID:7718557
A/Accession: A56318
A/Molecule type: mRNA
A/Residues: 1-1019 <KIT>
A/Cross-references: GB:U09860; NID:9746412; PIDN:AA050138.1; PID:G746413
R/Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A/Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A/Reference number: A43090; MUID:94329561; PMID:8052624
A/Accession: B43090
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 749-1019 <KIT2>
A/Cross-references: GB:U09860
C/Comment: The mechanism of association with the membrane of the intestinal brush border
ctated below) or with amino-terminal myristoylation of the heavy chain.
A/Gene: GDB:PRSS7
A/Cross-references: GDB:394083; OMIM:226200
A/Map position: 21q21-21q21
C/Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)

ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms invol
ducts.
C/Function:
A/Description: cleaves activation peptide from trypsinogen to produce active trypsin
A/Pathway: intestinal digestive hydrolase cascade
C/Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding r
C/Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:1-784/Product: enteropeptidase heavy chain #status predicted <HC>
F:22-38/Domain: transmembrane #status predicted <TM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: NMM homology <MM>
F:526-631/Domain: C1r/C1s repeat homology <C1r>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SR
F:785-1019/Product: enteropeptidase light chain #status predicted <LC>
F:785-1014/Domain: trypsin homology <TRY>
F:116,147,179,328,335,388,440,470,503,630,682,706,725,848,887,909,949/Binding site
F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 23.8%; Score 647; DB 1; Length 1019;
Best Local Similarity 35.8%; Pred. No. 6.8e-39;
Matches 138; Conservative 67; Mismatches 154; Indels 26; Gaps 10;

QY 113 CNSGIECDSSGTCINPNSWCDGVSHCPGGEDENRCVRLYG-----ENFLOVYSSQ 164
Db 643 CHADHFQC-KNGECVPLVNLCDGLHKEGSDGDEADCVRFNGTNNGLVRFRIQ----- 696
QY 165 RKSHPVQDDNNENYGRAACRDGMYKNNFYSSQGIYDDSGTSFMKLNLSAGNVDIYK 224
Db 697 -SIWHTACAENWTTQISNDVQCLGLGSG-NSSKPIFTSDGG-PFVKLNTAP---DGHLLI 750
QY 225 LVHSDACSSKSAVVSURC--IACGVNLSRS--QSRIVGGSALPGAMPQVSLHVQNVHVC 281
Db 751 LPSQOCLQDSILRLQCNHKSCKGLAAQDITPKLVGSKAKGAWPVVGLYGGELLC 810
QY 282 GSGIITPWIIVTAACHVEKPLNNPHWTFAPAGILRQSFME--YGAGVQVEKVISHPNYDS 339
Db 811 GASLVSSDMLVSAACHVYGRNLEFSKWTAILGLHMKNLATSPQVPLDIIEVINPHYNR 870
QY 340 KTKNDIALMKLQKLTENDLVKVPCLPNPGMLQPLQCLWISGWGATEKGTSEVLYNA 399
Db 871 REKNDIAMHLEKFNVTYDIQICLPEENQVPPGRCISAGWALYVGGTITANILQE 930
QY 400 AKVLLIETQCNRSRVYDNLITPAMICAGFLQGVNDSCQDSGGPLVTSKNNIWLIGDT 459
Db 931 ADVPLLSNERCQQQMPEYN--ITENMVCAGYEAGGVDSQCGDSGGPLMCCQNNRFLAGVT 989
QY 460 SMGSGCAKAYRPGVYGVNVVFTDWI 484
Db 990 SFGYKCALPNRPGVYARVRFTEWI 1014

RESULT 3
A53663
enteropeptidase (EC 3.4.21.9) precursor [validated] - pig
N/Alternate names: enterokinase
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C/Accession: A53663
R/Watsushima, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa,
J. Biol. Chem. 269, 19976-19982, 1994
A/Title: Structural characterization of porcine enteropeptidase.
A/Reference number: A53663; MUID:94327548; PMID:8051081
A/Accession: A53663
A/Molecule type: mRNA
A/Residues: 1-1034 <MAT>
A/Cross-references: GB:D30799; NID:G505122; PIDN:BA06459.1; PID:G505123
A/Note: parts of this sequence, including the amino ends of three chains isolated from
C/Comment: The mechanism of association with the membrane of the intestinal brush bord
ctated below) or with amino-terminal myristoylation of the heavy chain.
C/Complex: Mature enteropeptidase is variously reported to contain two (heavy and light
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms invol

C;Function: cleaves activation peptide from trypsinogen to produce active trypsin
A;Pathway: intestinal digestive hydrolase cascade
C;Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding repeat
C;Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F;22-38/Domain: transmembrane #status predicted <TM>
F;52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F;118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
F;119-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;357-519/Domain: MAM homology <MAM>
F;541-646/Domain: C1r/C1s repeat homology <C1R>
F;658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC>
F;800-1034/Product: enteropeptidase light chain #status predicted <LCH>
F;800-1029/Domain: trypsin homology <TRY>
F;116,147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,96
F;787-911,825-841,925-982,956-971,982-1010/Disulfide bonds: #status predicted
F;840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 23.3%; Score 634; DB 1; Length 1034;
Best Local Similarity 33.3%; Pred. No. 6.1e-38;
Matches 147; Conservative 72; Mismatches 179; Indels 44; Gaps 13;

QY 65 VCTQPSGTVCTSKKALCIT---ITLG-----TFLVGAALAGLWKGKCSNSG 117
DB 610 VYTGPGVDFVFTTRMTVLTINDALTGKGFKANFTGYHLG-----IPEPCKEDN 662
QY 118 IECDSSGTCINPSNWCDSVSHCPGGEDENRCVRLYG--PNFTIQVYSSQKSWHPVCDD 175
DB 663 FQCE-NGECVLLVNLCDGFSHCKDGSDEAHCVRFLNGTANNGLVQFRIQSLWHTACAA 721
QY 176 WNEYGEACRDGMKYN-----FYSSQGIIVDSGSTSPMKLNTSAGNVDIYKLIYHSDA 230
DB 722 WTTQTSDDVQCLGLGTGNSMPFFS-----SGGGPFVKLNTAPNGSLI---LTASEQ 771
QY 231 CSSKAVVSLRC--IAGCVNLNSRQS-RIVGGESALPGAMPQVSLHVQNVHVCSSIT 287
DB 772 CFEDSLILLCNKHCKGKQVAGVSPKIVGNDREGAPVWVALYNGQLLOGASIVS 831
QY 288 PEMIVTAACHVCEPLNPNPHWTAFAGILRQSPFMFYAGVQ-----VEKVIHNPVDSKTK 342
DB 832 RDLVLSAAHCVCYGRNLEPESKKAIGLHMTSL---TSPQIVTRLIDIDINPHNRRK 888
QY 343 NNDIALMKLQKPLTFNDLVKPCVLPNPGMLOPEQLCMTSGWGAEEKGTSEVLNAAKV 402
DB 889 DSDIAMHLEFKVNYTDYIOPICLPENQVFPFGRICTSIAGKVIYQGSFADILQEA 948
QY 403 LLETQRCNSRYVDNLITPAMICAGFLQGNVDSGCGSGGLPVTSKNINWLIQDTSWG 462
DB 949 PLUSNEKCCQOQMEPEYN-ITENWMCAGYEGGIDSCQSGSGGLPMTCLNNRLLAGVTSFG 1007
QY 463 SGCAKAYRPGVYGNVMVFTDWI 484
DB 1008 YQCALPNREGVYARVPKFTWI 1029

RESULT 4
JB0315
low-density lipoprotein receptor-related protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Feb-2003
A;Accession: JE0315
R;Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.
J. Biochem. 124, 784-789, 1998
A;Title: A novel low-density lipoprotein receptor-related protein with type II membrane
A;Reference number: JB0315; MUID:98429596; PMID:9756624
A;Accession: JE0315
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1113 <TM>
A;Cross-references: DBJ:AB013874; NID:93869144; PIDN:BA034371.1; PID:93869145
C;Superfamily: mouse low-density lipoprotein receptor-related protein; LDL receptor liga
F;337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F;374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;648-682/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F;723-757/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F;723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F;869-1097/Domain: trypsin homology <TRY>

Query Match 21.3%; Score 577.5; DB 2; Length 1113;
Best Local Similarity 33.3%; Pred. No. 8.1e-34;
Matches 134; Conservative 67; Mismatches 155; Indels 47; Gaps 14;

QY 111 SKCSNGIECDSSGTCINPSNWCDSVSHCPGGEDENRCVRLY---GNFTIQVYSSQKKS 167
DB 721 SFQDQNELAC-ANHECVPELDACDGVDCSDSSDSDGVCVTLKNGNSSLLTVHKSAG- 778
QY 168 WHPVCOQDNNYGRACDMGYKNFYSSQGIIVDSGSTSPMKLNTSAGNVD---IYKK 224
DB 779 -HRVCADGWRETLSQLACKQMGLEP--SVTKLIFGQEQQWRLYRYPWENLSTLOEL 835
QY 225 LVHSDACSSKAVVSLRCIA--CGVNLNSRQSRIVGGESALPGAMPQVSLHVQ-NVHVC 281
DB 836 LYVHSCPSERSISLLCSKQDCGRPPAARMKRIILGRTSRPGRWPWQCSLQSEPSGHIC 895
QY 282 GSGIITPEWIVTAACHVE-----KPLNPNPHWTAFAGILRQSPFMFYAGVQ 327
DB 896 GCVLIAKKVLTVAHCFEGREDADYKVVFGINNLDHP-----SGFMQTRF----- 941
QY 328 VEKVIHNPVDSKTKNDIALMKLQKPLTFNDLVKPCVLPNPGMLOPEQLCMTSGWGA 387
DB 942 VKTILLHPYSAVVDYDISVVELSDDINETSYRVPCLPSPPEVLEPDTYCYITGNG-- 999
QY 388 BEKGTSEVLNAAKVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGCGSGGLPVT 447
DB 1000 HMGKNKPFLLQGEVRIIPLEQCS-YFDMKTITNRMICAGVESGTVDSCMGDSGGLVC 1058
QY 448 SK-NNIIWLLIGTSGWGGC-AKAYRPGVYGNVMVFTDWIYRQM 488
DB 1059 ERPGGQWTLFGITSGWSVCFSKVLGPGVYNSVYFVGNIERQI 1101

RESULT 5

K0HUP

plasma kallikrein (EC 3.4.21.34) precursor - human
N;Alternate names: kininogenin; plasma prekallikrein
C;Species: Homo sapiens (man)
C;Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
A;Accession: A00921; A37939
R;Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A;Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four t
A;Reference number: A00921; MUID:86243359; PMID:3521732
A;Accession: A00921
A;Molecule type: mRNA
A;Residues: 1-638 <CH>
A;Cross-references: GB:M13143; NID:9190262; PIDN:AAA60153.1; PID:G190263
R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A;Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of
A;Reference number: A37939; MUID:91152016; PMID:1998666
A;Accession: A37939
A;Molecule type: protein
A;Residues: 20-27740-46; 'X', 48, 'H', 50, 'X', 52-70, 'H', 75-76, 'X', 78-80, 103-113, 131-140, 141-
1260-483, 'X', 285, 287-291, 'X', 293-295, 314-317, 'X', 319-320, 321-324, 'X', 329-333, 334-339, 'X',
525, 538-551, 562, 'X', 564-567, 573, 'X', 575-576, 578-583, 'X', 585, 592-604 <MCM>
C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C;Comment: The zymogen is activated by factor XIa, which cleaves the molecule into a 11
are linked by one or more disulfide bonds.
C;Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal r
inogen and may also play a role in the renin-angiotensin system by converting prorenin i
C;Genetics:
A;Gene: GDB:KLK3
A;Cross-references: GDB:127575; OMIM:229000

A:Map position: 4935-4935
C:Superfamily: coagulation factor XI, trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-638/Product: plasma kallikrein #status predicted <NAT>
F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-621/Domain: trypsin homology <TRY>
F:391-621/Domain: trypsin homology light chain #status predicted <LCH>
F:21-104,47-77,51-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,363
F:127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:318-347,340-345/Disulfide bonds: #status predicted
F:390-391/Cleavage site: Arg-1le (coagulation factor XIa) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.0%; Score 569.5; DB 1; Length 638;
Best Local Similarity 32.8%; Pred. No. 1.6e-33;
Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;
QY 55 RVLQASNPVCTQPSGTVCTSKKALCITLTGLFVGAALAAAGLLWKMGWS--- 111
DB 218 RVLQ--PDAPVCR-----TICTHYPN---CLPTFFYT-----NWKLESQNV 255
QY 112 ---KCSNGIECDG-----SG-----TCINP-----SNWCDGVSHCPGGEDENRVRL 151
DB 256 CLKLTSEGTSPSSSTPQENTISGYSLTCKRTLPCHSKIPYGVDF--GGEELN----- 308
QY 152 YGFNFILQVYSQRKSWHPVCDQDNNYNGRAACDMGKNNFYSSQGIQVDSGSTS--- 208
DB 309 --VTFVKGV-----NVQDS-----TCYK--IRCOFFYSLLPDCKECKC 347
QY 209 FMKLTASAGNVDTYKLYHSIDACSSKAVVSLRCIACGVN--LNSRSQRIRVGGESALPGA 266
DB 348 FLRLSDNGSPTRI---AYGTQSSG---YSLRLCNTGNSVCTTKTSTRIYVGTGNSWGE 401
QY 267 WPMQVSLHQV---NVHVCVGSIIIPWIVTAACHVE-KPLNPNWHTAFAGILRQSFMY 322
DB 402 WPMQVSLVQLTAQRLHGLGSLIGHQWLVTAACHDGLPLQDVR--IYSGILSLDITK 459
QY 323 GAGY-QUEKVIHPNVDSYKNDIALMKQLPFLNDLVKPVCLPNFGMMLQPEQLCWI 381
DB 460 DTPFSQIKETIIHQNTKVSSEGNHDIKLOAPLNNTYEFQKICLPKSGDSTTYTNGW 519
QY 382 SGWGAEEKGTSEVLNNAKVLIIETQRCNSRYVDNLITPAMICAGFLQGNVDSQGD 441
DB 520 TGRGFSKKEGEIQNILOKYNIPLVNBECKEY-QDYKITQPMVCAGYKEGKCKACKDS 578
QY 442 GGLVTSKNNIWLIGDTSWGGCAKAYRPGYVGNMVFDTWYIQMR-ADG 492
DB 579 GGLVCKNGHGMRLVIGTSWGGCAKARREGQYVTKVAYMYDWILEKTSQSDG 630

RESULT 6
S00845
hepsin (EC 3.4.21.-) - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1998 #sequence_revision 31-Dec-1998 #text_change 18-Jun-1999
C:Accession: S00845
R:Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dom
A:Reference number: S00845; PMID:88209431; PMID:2835076
A:Accession: S00845
A:Molecule type: mRNA
A:Residues: 1-417 <LEY>
A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064
C:Genetics:
A:Gene: GDB:HPN; TPRS51; hepsin
A:Cross-references: GDB:135685; OMIM:142440
A:Map position: 19q11-19q13.2

C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:23-45/Domain: transmembrane #status predicted <TM>
F:163-400/Domain: trypsin homology <TRY>
F:188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted
F:203,257,353/Active site: His, Asp, Ser #status predicted
Query Match 20.8%; Score 565.5; DB 1; Length 417;
Best Local Similarity 30.7%; Pred. No. 1.9e-33;
Matches 133; Conservative 68; Mismatches 173; Indels 59; Gaps 9;
QY 77 CTSKTKKALCITLTGLFVGAALAAAGLLWKMGSKCSNGSIECDSSGTCINPNWCDGV 136
DB 12 CCSRPKVA---ALTAGTLLTLLTAICA-----ASMAIVA 41
QY 137 SHCPGGEDENRVLYGPNFLOVYSSQRKSWHPVCDQDNNYNGRAACDMGKNNFY 196
DB 42 VLLRSDQELPYVQVSSADARLMVDFKTEGTWRLLCSSRNARVAGLSCEMGFLRALTH 101
QY 197 SGQIVDDSGSTSMKLTNSAGNVDIYK-KLYHSDA-----CSSKAVVSLRCIACGV 246
DB 102 SELDVRTAGA-----NGTSGFFCVDEGRPLHQLLELVISVCDPGRFLAALCQDGG- 154
QY 247 NLNRSQRIRVGGESALPGAAMPQVSLHVNQVHVCVGSIIIPWIVTAACHVEKPLNPNW 306
DB 155 -RRKLPVTRIVGRTSLGRWPMQVSLRYDGAHLGGSLLSGDWLTAACHCFPERNVLS 213
QY 307 HWTAFAGILROSFMFYGAGYQVEKVIHPNY-----DSKTNNDIALMKQLPFLNDL 360
DB 214 RWRVFAVAGVQASP-HGLQLGVQVAVYHGGYLPFRDPNSENSENDIALVHLSLPLTEY 272
QY 361 VKPVCLPNFGMMLQPEQLCWSIGWGAETEEKTSEVLNNAKVLIIETQRCNSRYVDNL 420
DB 273 IQPVCPLPANGQALVDKICTVGTGNTQYVQAGVQLQEARVPIISNDVNGADFYGNQI 332
QY 421 TPAMI CAGFLQGNVDSQGDSSGGLV-----TSKNNIWLIGDTSWGGCAKAYRPGYGN 476
DB 333 KPMFCAGYPEGGIDACQDSSGPPFCVDSISRTPRMLCGIVSWGTGTCALQKPGYTK 392
QY 477 VMVFTDWIYQMR 489
DB 393 VSDFFRWIFQAIK 405

RESULT 7
JC7731
membrane-bound arginine-specific serine proteinase precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 03-Feb-2003
C:Accession: JC7731; JC7775
R:Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Achauda,
J. Biochem. 130, 425-430, 2001
A:Title: Characterization of a membrane-bound arginine-specific serine protease from ra
A:Reference number: JC7731; PMID:21421307; PMID:11530019
A:Accession: JC7731
A:Molecule type: mRNA
A:Residues: 1-855 <KIS>
A:Cross-references: DBJ:AB049189
A:Experimental source: strain Male, 7-week-old
R:Satom, S.; Yamazaki, Y.; Tsuzuki, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.
Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
A:Title: A role for membrane-type serine protease (MT-Sp1) in intestinal epithelial tur
A:Reference number: JC7775; PMID:11573963
A:Content: Small intestine
A:Accession: JC7775
A:Molecule type: mRNA
A:Residues: 1-855 <SAT>
A:Cross-references: DBJ:AB037898
C:Comment: This enzyme, an epithelial-derived, type II integral membrane serine proteas
of specific proteins or peptides on the brushborder membranes. It also participates in
lial migration and/or cell loss.
C:Genetics:
A:Gene: mt-sp1

A;Map position: basolateral cell surface
A;Superfamily: membrane-bound arginine-specific serine proteinase
C;Keywords: protein digestion

Query Match 20.8%; Score 564.5; DB 2; Length 855;
Best Local Similarity 33.0%; Pred. No. 5.2e-33;
Matches 128; Conservative 62; Mismatches 131; Indels 67; Gaps 11;

QY 110 GSKGNSGIECDSSGTCINPNWCDGSHGFCGGEDENRCVRLYGNFILQVYSSQKRSWH 169
DB 522 GCSCPAGSFKC-SNGKCLPQSQOQNGKDDCGGDEASCDNVNAVSTCKYTRCQ-----N 576

QY 170 PVOQDDWN-ENYGRAACDMGMYKXNFYSSQGIIVDDSGSTSPMKLNTAGNVDIYKLYHS 228
DB 577 GLCLNKGNEPCDKGKDCSDGDEKNC-----DCGLRSFTK----- 611

QY 229 DACSKAVVSLRCIACGVNLNRSRQIRIVGGESALPGAPWQVSLH-VQNVHVCGSGIIT 287
DB 612 -----QARVVGNTNDEGEWPNQVSLHALGQHLGCABLIS 647

QY 288 PEWIVTAACHVEK-----FLNNPFWHTAFAGILRQSFYGYAGYQ-----VEKVISHPNYDSK 340
DB 648 PDWLVSAAHCFODEIFKYSDHTMTWTAFLGLDQS-KXSASGVQEHKLRITHPSPNDF 706

QY 341 TKNDIALMKLOKPLTFNDLVKPVCLPNPMMLOPEQLCWISGSGATEEKGKTEVLNAA 400
DB 707 TFDYDIALLELEKPAEYSTVVRPCLPDNTHVFPAGKAIWVTGHTKEGGTGALILQKG 766

QY 401 KVLLETQKNSRYVYDNLITPAMICAGFLQNVDSQCGSGPLVT-SKNNIWLIGDT 459
DB 767 EIRVINQTCBE--LLPQOITPRMVCVFLSGVDSCQDGGPLSSVEKDGRIPOAGVV 824

QY 460 SNGSGCAKAYRPGVYGNVNFDTWYRQ 487
DB 825 SNGEGCAQRNKGVYTRIPEDVDWIKEQ 852

RESULT 8
KQRTPL
plasma kallikrein (EC 3.4.21.34) precursor - rat
N;Alternate names: Fletcher factor; kininogenin; serum kallikrein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 18-Jun-1999
C;Accession: A39180; A33320; S06851; I53041; S06852
R;Seaidah, N.G.; Rosinski-Chupin, I.; Mattei, M.G.; Mbikay, M.; Chretien, M.; Seidah, N.G.
Biochemistry 30, 1628-1635, 1991
A;Title: Gene structure and chromosomal localization of plasma kallikrein.
A;Reference number: A39180; MUID:91129236; PMID:1993180
A;Accession: A39180
A;Molecule type: DNA
A;Residues: 1-638 <SEA>
A;Cross-references: GB:J05315
A;Note: the authors translated the codon GAG for residue 81 as Gln
R;Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazur
DNA 8, 563-574, 1989
A;Title: The cDNA structure of rat plasma kallikrein.
A;Reference number: A33320; MUID:90091743; PMID:2598771
A;Accession: A33320
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-638 <SEI>
A;Cross-references: GB:M30282; NID:G205010; PIDN:AAA41463.1; PID:G205011
A;Note: part of this sequence, including the amino ends of both the heavy and light chain
R;Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
Biochim. Biophys. Acta 999, 103-110, 1999
A;Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
A;Reference number: S06851; MUID:90089457; PMID:2597701
A;Accession: S06851
A;Molecule type: protein
A;Residues: 20-45;391-413 <PAQ>
R;Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazur
DNA Cell Biol. 8, 563-574, 1989
A;Title: The cDNA structure of rat plasma kallikrein.

A;Reference number: I53041
A;Accession: I53041
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-638 <RES>
A;Cross-references: GB:M58590; NID:G206721; PIDN:AAA42069.1; PID:G206722
C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex
C;Comment: The zymogen is activated by factor XIIIa, which cleaves the molecule into a 1
are linked by one or more disulfide bonds.

C;Genetics:
A;Gene: PK
A;Superfamily: coagulation factor XI; trypsin homology
C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infl
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-390/Product: plasma kallikrein heavy chain #status experimental <MAR1>
F;20-109/Domain: apple repeat <AP1>
F;110-199/Domain: apple repeat <AP2>
F;200-289/Domain: apple repeat <AP3>
F;291-380/Domain: apple repeat <AP4>
F;391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F;391-621/Domain: trypsin homology <TR1>
F;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,32:
F;127,215,308,453,459,494/Binding site: carboxydrate (Asn) (covalent) #status predicted
F;396/Binding site: carboxydrate (Asn) (covalent) #status experimental
F;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 20.5%; Score 556; DB 1; Length 638;
Best Local Similarity 36.8%; Pred. No. 1.5e-32;
Matches 117; Conservative 59; Mismatches 128; Indels 16; Gaps 7;

QY 181 GRAACDMGYKN---NFYSSQGIIVDD---SGSTSPMKLNTAGNVDIYKLYHSDACSSK 234
DB 314 GADACQCTCTKTRCQFFTVSLLPQDCAEGCKSLRLSTDGSPTRITVEAQSSGYSLR 373

QY 235 AVVSLRCIACGVNLNRSRQIRIVGGESALPGAPWQVSLHVNQV---HVCVGGSIITPEWI 291
DB 374 LCKVRESSDCTTKIN---ARIVGNTSSLGEMWQVSLQVCLVSNHMCSSGIIQRQWI 429

QY 292 VTAACHVEKPLNPNFWHTAFAGILRQSFYGYAGY-QVEKVISHPNYDSKTKNDIALMK 350
DB 430 LTAACHFD-GIPYDVWRIYVGGILNLSEITNKTFFSSIKELITHQYKMSGSEGYDIALIK 488

QY 351 LQKPLTFNDLVKPVCLPNPMMLOPEQLCWISGSGATEEKGKTEVLNAAKVLIIETORC 410
DB 489 LQTLFNTYEFQKPICLFSKADTNTIYTCNVGTGWTGKERTQNTIQLKATIPLVNNEEC 548

QY 411 NSRYVDNLITPAMICAGFLQNVDSQCGSGPLVTSKNNIWLIGDTSWGGCAKAYR 470
DB 549 QKYK-RDYVITKQVICAGYKEGGIDACKDGGSGPLVCKHSGRWLVGITSWGGCAKKEQ 607

QY 471 PGYVGNVWFTDWIRQMA 490
DB 608 PGVYTKVAEYIDWILSKIQS 627

RESULT 9
KQMSPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 18-Jun-1999
C;Accession: A36557
R;Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemor
DNA Cell Biol. 9, 737-748, 1990
A;Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparis
A;Reference number: A36557; MUID:91090844; PMID:2264928
A;Accession: A36557
A;Molecule type: mRNA
A;Residues: 1-638 <SEI>
A;Cross-references: GB:M58588; NID:G200358; PIDN:AAA63393.1; PID:G200359
A;Note: part of this sequence, including the amino ends of both the heavy and light chain
C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex
C;Comment: The zymogen is activated by factor XIIIa, which cleaves the molecule into a 1
are linked by one or more disulfide bonds.

1

A;Reference number: S33777; MUID:93305733; PMID:8318546

A;Accession: S33777

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-416 <PAR>

A;Cross-references: EMBL:X70900; NID:g57928; PIDN:CAAS0256.1; PID:g57929

C;Superfamily: hepsin; trypsin homology

C;Keywords: hydrolase; liver; serine proteinase; transmembrane protein

F;22-44/Domain: transmembrane #status predicted <TMN>

F;162-399/Domain: trypsin homology <TRY>

F;187-203,290-358,321-337,348-380/Diulfide bonds: #status predicted

F;202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 19.5%; Score 529.5; DB 1; Length 416;

Best Local Similarity 30.6%; Pred. No. 7.6e-31;

Matches 132; Conservative 66; Mismatches 177; Indels 57; Gaps 10;

Qy 77 CTSKTKKALCITITLGTFLVGAALAAAGLLWKFMSKCSNSGTECDSSGTCINPSNWCQGV 136

Db 11 CCSRPKVA---ALTVGTL-----FL-----TGGAASWAIVT 40

Qy 137 SHCPGGEDENRCVRLGPNFILOYSSQKSWHPVQDDWNYGAAACRDMGYKNFY 196

Db 41 ILRSDOEPLYQVQLSGDSRLVLDKTEGTRWLLCSRSNARVAGLCEENGFLRALAH 100

Qy 197 SQGIVDDSGSTSFMKLNTSA-----GNVDIVYKLYHSDA---CSSKAVWSLRCTACGVN 247

Db 101 SELDVRTAGAN-----GTSGFVCDGGLPLAQLRLDVISVCDPCRGFLATQCDCG-- 153

Qy 248 LNSRSRRIYGGESALPGAPWQVSLHVQVHVVHVGSGSIITPWIVTAHCEKPLNPNWH 307

Db 154 RRLPVDRIYGGQDSSLGRMPQVSLRYDGLHCGSLSGDWLVTAACHCFERNRVLRS 213

Qy 308 WTAFAGLRQSFYFAGYQVERVISHPNY-----DSKTKNDIALMKLQKPLTFNDLV 361

Db 214 WRVFAVAVATSP-HAVOLGVQAVIYHGGYLPFPDPTDENSNDIALVHSSLPLEYI 272

Qy 362 KPVCPLNPGMLOPQELCWISGWGATEEKGKTSVNLNAKULLIETQCNRYVDNLIIT 421

Db 273 QPVCPLPAAGALVDGKVCVTGTGNGTQFYGOQAVLQEARVPIISNEVCNPFDFYGNQIK 332

Qy 422 PAMICAGFLQGNVDSQQGSGGGLV---TSKNNTWMLIGTSGSGCAKAYRPGVYGNV 477

Db 333 PKMFCAGYPEGIDACQSGSGHFCEDRISGTSRRLCGIVSGWTGALARKFGVYTKV 392

Qy 478 MVFTDIYRQMR 489

Db 393 IDPREMIFQAIK 404

RESULT 12

PLMS

Plasmin (EC 3.4.21.7) precursor - mouse

N;Contains: angiotensin; plasminogen

C;Species: Mus musculus (house mouse)

C;Date: 20-Sep-1991 #sequence revision 01-Nov-1996 #text_change 18-Jun-1999

C;Accession: A38514; S48202; S48203

R;Degen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.

Genomics 8, 49-61, 1990

A;Title: Characterization of the cDNA coding for mouse plasminogen and localization of

A;Reference number: A38514; MUID:91184812; PMID:2081600

A;Accession: A38514

A;Molecule type: mRNA

A;Residues: 1-812 <DEG>

A;Cross-references: GB:J04766; NID:g200402; PIDN:AAA50168.1; PID:g200403

R;Liijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.

Eur. J. Biochem. 224, 863-871, 1994

A;Title: Characterization of the murine plasma fibrinolytic system.

A;Reference number: S48202; MUID:95010076; PMID:7521120

A;Accession: S48202

A;Molecule type: protein

A;Residues: 20-25 <LIJ>

A;Accession: S48203

A;Molecule type: protein

A;Residues: 22-27 <LIJ>

C;Comment: Plasminogen is synthesized by the kidney and is present in plasma and many o
mediately after dissociation from the clot. In the presence of the inhibitor, the activ
e inhibitor, the activation involves also removal of the activation peptide.

C;Comment: Stromelysin 1 (see PIR:KCMSS1) acts on plasminogen to produce angiotatin. T
eul in treating solid tumors.

C;Function:

A;Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a v
ns the walls of the graafian follicle; also activates the urokinase-type plasminogen ac

A;Pathway: fibrinolysis

C;Superfamily: plasmin; kringe homology; plasminogen-related protein precursor homology

C;Keywords: angiotensin inhibitor; blood; duplication; fibrinolysis; glycoprotein; hyd

F;1-96/Domain: plasminogen-related protein precursor homology <PLPH>

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-812/Product: plasminogen #status predicted <PRO>

F;20-96/Domain: activation peptide #status predicted <APT>

F;79-466/Product: angiotensin #status predicted <AST>

F;97-581,582-812/Product: plasmin #status predicted <MAT>

F;97-581/Domain: chain A #status predicted <ACH>

F;103-181/Domain: kringe homology <KR1>

F;185-262/Domain: kringe homology <KR2>

F;275-352/Domain: kringe homology <KR3>

F;377-454/Domain: kringe homology <KR4>

F;481-560/Domain: kringe homology <KR5>

F;582-812/Domain: chain B #status predicted <BCH>

F;582-805/Domain: trypsin homology <TRY>

F;49-75,53-63,103-181,124-164,152-176,185-262,188-316,206-245,234-257,275-352,296-335,3

Bonds: #status predicted

F;79-79/Cleavage site: Glu-Asn (stromelysin 1) #status predicted

F;136,308/Binding site: carboxylate (Asn) (covalent) #status predicted

F;466-467/Cleavage site: Thr-Val (stromelysin 1) #status predicted

F;581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental

F;624,667,763/Active site: His, Asp, Ser #status predicted

Query Match 19.1%; Score 518.5; DB 1; Length 812;

Best Local Similarity 32.7%; Pred. No. 1e-29;

Matches 136; Conservative 54; Mismatches 143; Indels 83; Gaps 16;

Qy 106 WKFMGSK-CSNSGIECDSSGTCI-----NPSNWCQGVSHC--PGGEDENRCVRLYXPN 155

Db 446 WEVCNLRKCSGT-----GSVVELPTVSGFSGFSDSDTCMYGKDYRGKTAVTAAG 499

Qy 156 FLQVYSSQKSWHPVC-----QDDWNYGAAACRDMG-----YKKNFYSSQGIYVDDSG 205

Db 500 TPCQGWAAQEPHRSIFTPQTNPRADLEKNYCRNPDGVDVNGPWCYTTN----- 547

Qy 206 STSPMKLNTSAGNVDIYKLYHSDACSKSAVUSLRCTACGVNLNLSRQ--SRIVGGESAL 263

Db 548 -----PRKLY--DYCDIFLCASASSFECKGKQVPEKKCFGRVVGCVAN 589

Qy 264 PGANPWQVSLHVQ--NVHVCSSIIITPEWIVTAHCEKPLNPNWHWTAFAGILRQSFMF 321

Db 590 PHSWQVSLRTRFTGQHFQCGTLIAPEWLVTAACHLEKS--SRPEFYKVLGAHEE---- 644

Qy 322 YGAGYQVR-----KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPNFGMLQPE 376

Db 645 YIRGLDVQVSEISVAKLLEPN-----NRDIALKLSRPATITDKVIPACLPSPNYWVADR 698

Qy 377 QLCWISGWGATE--EKGKTSVNLNAKULLIETQCNRSRVVDNLTIPAMICAGFLQGN 433

Db 699 TICITIGMETQGTFGAGRLKE---ACLPIENKVCNRYEYLNVRVKSTELCAGQLAGG 754

Qy 434 VDSQCGDSGGFLVTSKKNINWMLIGTSGSGCAKAYRPGVYGNVWFTDIYRQMR 489

Db 755 VDSQCGDSGGFLVCFEKDKYILQGVTSWGLGCAKPNKPGVYVVRVSRFVDWIEMR 810

RESULT 13

S11674

acrosin (EC 3.4.21.10) precursor - human

C;Species: Homo sapiens (man)

C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999
 C/Accession: A34170; S08994; S02428; S04940; S16657; S02780; S10695; S12968
 R/Baba T.; Kashiwabara, S.; Watanabe, K.; Itoh, H.; Michikawa, Y.; Kimura, K.; Takada, J. Biol. Chem. 264, 11920-11927, 1989
 A>Title: Activation and maturation mechanisms of boar acrosin zymogen based on the deduced amino acid sequence
 A/Reference number: A34170; MUID:89308595; PMID:2745422
 A/Accession: A34170
 A>Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-415 <BB>
 A/Cross-references: GB:J04950; NID:G164702; PIDN:AAA31131.1; PID:G164703
 R/Cechova, D.; Toepfer-Petersen, E.; Zucker, A.; Jonakova, V. Biol. Chem. Hoppe-Seyler 371, 317-323, 1990
 A>Title: Is sperminogen a modified proacrosin? Isolation, purification, and partial characterization of boar acrosin
 A/Reference number: S08994; MUID:90253655; PMID:2111146
 A/Accession: S08994
 A/Molecule type: protein
 A/Residues: X, 18, X, 20-25, 'X', 27-32, 'X', 34-38, 'X', 40-50 <CEC>
 R/Toepfer-Petersen, E.; Henschen, A. FEBS Lett. 226, 38-42, 1987
 A>Title: Acrosin shows zona and fucose binding, novel properties for a serine proteinase
 A/Reference number: S02428; MUID:88083633; PMID:3480243
 A/Accession: S02428
 A/Molecule type: protein
 A/Residues: 17-32/40-55 <TOE>
 R/Adham, I.M.; Klemm, U.; Mater, W.M.; Hoyer-Fender, S.; Tsaousidou, S.; Engel, W. Eur. J. Biochem. 182, 563-568, 1989
 A>Title: Molecular cloning of preproacrosin and analysis of its expression pattern in spermatogenesis
 A/Reference number: S04940; MUID:89325301; PMID:2502391
 A/Accession: S04940
 A/Molecule type: mRNA
 A/Residues: 1-7,9-210, 'Q', 212-216, 'VT', 219-346, 'A', 348-388, 390-393, 'GN', 396, 'LVE', 399-400
 A/Cross-references: EMBL:X14844
 A/Note: the authors translated the codon CCT for residue 240 as Ala, GCC for residue 264
 R/Adham, I.M. submitted to the EMBL Data Library, March 1989
 A/Reference number: S16657
 A/Accession: S16657
 A/Molecule type: mRNA
 A/Residues: 1-7,9-210, 'Q', 212-216, 'VT', 219-346, 'A', 348-388, 390-398, 'KELL', <AD2>
 A/Cross-references: EMBL:X14844; NID:G1867; PIDN:CAA32948.1; PID:G1868
 A/Note: the difference at the carboxyl end is due to a frameshift error
 R/Baba, T.; Michikawa, Y.; Kawakura, K.; Arai, Y. FEBS Lett. 244, 132-136, 1989
 A>Title: Activation of boar proacrosin is effected by processing at both N- and C-terminal
 A/Reference number: S02780; MUID:89171246; PMID:2494060
 A/Accession: S02780
 A/Molecule type: protein
 A/Residues: 17-69 <BA2>
 R/Toepfer-Petersen, E.; Steinberger, M.; von Eschenbach, C.E.; Zucker, A. FEBS Lett. 265, 51-54, 1990
 A>Title: Zona pellucida-binding of boar sperm acrosin is associated with the N-terminal
 A/Reference number: S10695; MUID:90306316; PMID:2365054
 A/Accession: S10695
 A/Molecule type: protein
 A/Residues: 40-62 <TO2>
 R/Toepfer-Petersen, E.; Calvete, J.; Schaefer, W.; Henschen, A. FEBS Lett. 275, 139-142, 1990
 A>Title: Complete localization of the disulfide bridges and glycosylation sites in boar acrosin
 A/Reference number: S12968; MUID:91085546; PMID:2261983
 A/Accession: S12968
 A/Molecule type: protein
 A/Residues: 17-29/34-66/68-91/94-121/123-166/171-184/190-207/209-216/219-228/231-245/248
 C/Superfamily: acrosin; trypsin homology
 C/Keywords: glycoprotein; hydrolase; serine proteinase; sperm
 F1-16/Domain: signal sequence #status predicted <SIG>
 F17-415/Product: acrosin #status experimental <NAT>
 F17-39/Product: acrosin light (A) chain #status experimental <LCH>
 F40-415/Product: acrosin heavy (B) chain #status experimental <HCH>
 F40-283/Domain: trypsin homology <TRY>
 F300-374/Region: proline-rich
 F19,208/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F22-152/26-160/71-87,175-244,207-223,234-264/Disulfide bonds: #status experimental

F:86,140,238/Active site: His, Asp, Ser #status predicted

Query Match 18.1%; Score 490.5; DS 1; Length 415;
 Best Local Similarity 37.7%; Pred. No. 5e-28;
 Matches 104; Conservative 41; Mismatches 78; Indels 53; Gaps 9;

QY 244 CGVNLSSRQS--RIVGGESALPGAWPCVSLHV-----QNVHYCGGSIITPEWILVTAA 295
 DB 26 CGLRFRQKLESQWRVVGNSAFPGAWPMVSLQIFWYHNNRRYHTCCGILLNSHWLITAA 85
 QY 296 HCVE-----KPLNNPWHWTAFAGILRQSFMYGAGYQVEKVIS 333
 DB 86 HCFKFKKKYTDRLFLFANEVWGSNKVPKP-----LQERF-----VEEII 128
 QY 334 HNYDSKTKNDIALMKLOKPLTFNDLVKPVCLP--NPGMMLOPQOLCWISGKATEEGK 391
 DB 129 HEKYVSGLEINDIALIKITPPVPCGPFIGFGCLPQFKAGPPRAP-QTCVWTGMYLKEKG 187
 QY 392 -KTSEVLNAAKVLLIETORCNSRYVVDNLIITPAMICAGFLQGNVDSQGDGSLV--TS 448
 DB 188 PRTPSLQEARVALIDLELCNSTRWYNGIRSTNVCAGYPRGKIDTCQGDGSLVLCRDR 247
 QY 449 KNNIWWLIGDTSGSGCAKAYRPGVYGVNVTDWI 484
 DB 248 AENTFVVVGITSGVGCARAKRPGVYTSTWPIYINWI 283

Search completed: June 1, 2004, 14:38:22
 Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 1, 2004, 16:37:28 ; Search time 688 Seconds
(without alignments)
3252.993 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 2717
Sequence: 1 MALNSGSPAIGPYENHGV.....VGVNVMVFDWIYRQNRADG 492

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09615285/runat_01062004_132135_20064/app_query.fasta_1.647
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09615285@cgn2_1_1.481 @runat_01062004_132135_20064
-NCPU=6 -ICPU=3 -NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
13:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
14:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
15:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
16:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
17:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
18:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
19:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

US-10-295-027-585
; Sequence 585, Application US/10295027
; Publication No. US2003023250A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Azi, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Ros Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666

1	2717	100.0	1479	16	US-10-295-027-585	Sequence 585, App
2	2712	99.8	2486	9	US-09-981-353-22	Sequence 22, Appl
3	2712	99.8	2486	13	US-10-116-802-203	Sequence 203, App
4	2711	99.8	3226	15	US-10-205-823-414	Sequence 414, App
5	2709	99.7	1476	9	US-09-759-143-931	Sequence 931, App
6	2709	99.7	1476	9	US-09-780-669-931	Sequence 931, App
7	2709	99.7	1476	9	US-09-822-827-931	Sequence 931, App
8	2709	99.7	1476	9	US-09-885-793-931	Sequence 931, App
9	2709	99.7	1476	9	US-09-895-814-931	Sequence 931, App
10	2709	99.7	1476	14	US-10-012-896-931	Sequence 931, App
11	2709	99.7	1476	15	US-10-144-678A-931	Sequence 931, App
12	2709	99.7	1476	15	US-10-294-025-931	Sequence 930, App
13	2709	99.7	1479	9	US-09-759-143-930	Sequence 930, App
14	2709	99.7	1479	9	US-09-780-669-930	Sequence 930, App
15	2709	99.7	1479	9	US-09-822-827-930	Sequence 930, App
16	2709	99.7	1479	9	US-09-895-814-930	Sequence 930, App
17	2709	99.7	1479	14	US-10-012-896-930	Sequence 930, App
18	2709	99.7	1479	15	US-10-144-678A-930	Sequence 930, App
19	2709	99.7	1479	15	US-10-294-025-930	Sequence 930, App
20	2709	99.7	1479	15	US-10-294-025-930	Sequence 929, App
21	2709	99.7	3245	9	US-09-759-143-929	Sequence 929, App
22	2709	99.7	3245	9	US-09-780-669-929	Sequence 929, App
23	2709	99.7	3245	9	US-09-822-827-929	Sequence 929, App
24	2709	99.7	3245	9	US-09-895-814-929	Sequence 929, App
25	2709	99.7	3245	9	US-10-012-896-929	Sequence 929, App
26	2709	99.7	3245	15	US-10-144-678A-929	Sequence 929, App
27	2709	99.7	3245	15	US-10-294-025-929	Sequence 929, Appli
28	2709	99.7	3245	15	US-10-334-038-5	Sequence 894, App
29	2704	99.5	2479	9	US-09-753-143-894	Sequence 894, App
30	2696	99.2	2479	9	US-09-780-669-894	Sequence 894, App
31	2696	99.2	2479	9	US-09-822-827-894	Sequence 894, App
32	2696	99.2	2479	9	US-09-895-814-894	Sequence 894, App
33	2696	99.2	2479	9	US-09-895-814-894	Sequence 894, App
34	2696	99.2	2479	10	US-09-776-191-69	Sequence 69, Appl
35	2696	99.2	2479	14	US-10-012-896-894	Sequence 894, App
36	2696	99.2	2479	15	US-10-144-678A-894	Sequence 894, App
37	2696	99.2	2479	15	US-10-294-025-894	Sequence 894, App
38	2696	99.2	2479	16	US-10-295-027-587	Sequence 587, App
39	2696	99.2	2479	16	US-10-295-027-587	Sequence 870, App
40	2696	99.2	2479	16	US-10-156-314A-36	Sequence 36, Appl
41	2696	99.2	2479	17	US-10-188-832-203	Sequence 203, App
42	2696	99.2	3298	15	US-10-235-699-6	Sequence 6, Appli
43	2116	77.9	1077	9	US-09-988-975A-2	Sequence 2, Appli
44	1569	57.7	1077	15	US-10-235-699-2	Sequence 2, Appli

; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US 60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1386
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 585
 ; LENGTH: 1479
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-295-027-585

Alignment Scores:
 Pred. No.: 3,186-307 Length: 1479
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 16 Gaps: 0

US-09-615-285b-2 (1-492) x US-10-295-027-585 (1-1479)

QY	1	MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
DB	1	ATGGCTTTGAATCAGGTCACCCAGCATTTGGACCTTACTATGAAAACCATGGATAC	60
QY	21	GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis	40
DB	61	CAACCGAANACCCCTATCCCGCACAGCCACTGTGTCCCTCTACGAGGTGCAT	120
QY	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
DB	121	CGGGCTCAGTACTACCCGTCCCGGTGCCAGTACGCCCGAGGGTCTCTGACGACGGCT	180
QY	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
DB	181	TCCAAACCCGTCGTCGACGAGCCCAATCCCCATCCGGGACAGTGTGCACCTCANAG	240
QY	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
DB	241	ACTAAGAAAGCACTGTGCATCACCTTGACCTGGGGACCTTCTCTGTGGAGCTGGCGCTG	300
QY	101	AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
DB	301	GGCGTGGCCCTACTCTGGAAGTTATGGCGAGCAAGTGTCCCACTCTGGGAATAGAGTGC	360
QY	121	AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro	140
DB	361	GACTCCTCAGGTACCTGCATCAACCCCTCAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	420
QY	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal	160
DB	421	GGCGGGAGGAGAGAAATCGGT	480
QY	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr	180
DB	481	TACTCATCTCAGAGGAAGTCTGGCACCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	540
QY	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200

RESULT 2
 US-09-981-353-22
 ; Sequence 22 Application US/09981353
 ; Patent No. US20020160392A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lasek, Amy W.
 ; APPLICANT: Jones, David A.
 ; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
 ; FILE REFERENCE: PA-0038 US
 ; CURRENT APPLICATION NUMBER: US/09/981,353

DB	541	GGGCGGGCGCTGCAGGACATGGGCTATAGAAATATTTTACTCTAGCCAGGAATA	600
QY	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
DB	601	GTGGATGACAGCGGATCCACGCTTTATGAACCAAGTCCGCGCAATGTCTGAT	660
QY	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
DB	661	ATCTATAAATACTGTACACAGTGTCTCTTCAAAAGCAGTGTCTTTTACGC	720
QY	241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu	260
DB	721	TGTATAGCTTGGGGGTCAACTTGAACCAAGCGGAGAGATCTGGCGGCGAG	780
QY	261	SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnValHisVal	280
DB	781	AGCGGCTTCCCGGGGCTTGGCCCTGCGAGTGCAGCTGCAGACGTCACAGTG	840
QY	281	CysGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys	300
DB	841	TGCGAGGCTCCATCATCCCCCGAGTGGATCGTACAGCCGCCCTGCGTGGAAAAA	900
QY	301	ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
DB	901	CCTCTTAACATCCATGCGCATTTGACGCGCATTTTGGGGATTTTGAGACATCTTTCA	960
QY	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys	340
DB	961	TTCTATGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATATGACTCCAAG	1020
QY	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
DB	1021	ACCAAGAACATGACATTTGCTGATGAAGCTGCAGAGCCTCTGACTTTTCAACGACCTA	1080
QY	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp	380
DB	1081	GTGAACACAGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGCAACAGCTCTCTGG	1140
QY	381	IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla	400
DB	1141	ATTTCCGGGTGGGGGCCCGCAGGAGAGAAAGGAGACCTCAGAGTGTGAACGCTGCC	1200
QY	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAsnLeuIle	420
DB	1201	AAGTGTCTTCTCATTTAGACACAGATGCAACAGAGATATGCTTATGACACCTGATC	1260
QY	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp	440
DB	1261	ACACAGCATGATCTGTGCGGCTTCTGCGAGGAAAGCTGATCTTTCAGGAGTAC	1320
QY	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpIleGlyAspThrSer	460
DB	1321	AGTGGAGGCTCTGTGTCTTCAAGCAATATCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1380
QY	461	TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe	480
DB	1381	TGGGTCTGTGCTGCCAAGCTTACAGACAGGAGTGTACGGGAATGTGTGTGTGTGTGTGT	1440
QY	481	ThrAspTrpIleTyrArgGlnMetArgAlaAspGly	492
DB	1441	ACGAGTGGATTTATCGACAAATGAGGCGACGCGC	1476

; CURRENT FILING DATE: 2001-10-11
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 22
 ; LENGTH: 2486
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. US20020160382A1 991163CB1
 US-09-981-353-22

Alignment Scores:
 Pred. No.: 2,68e-306 Length: 2486
 Score: 2712.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.82% Indels: 0
 DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-981-353-22 (1-2486)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 DB 149 ATGGCTTTGAATCAGGCTCAGCCCTGCTATTTGACCTTACTATGAAAAACCATGGATAC 208
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 DB 209 CAAACGGAAAAACCCCTATCCCGCAGACGCCACTGTGGTCCCCTGCTACGAGTGCTAT 268
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 DB 269 CCGGCTCAGTACTACCCGCTCCCGCTGCCAGTACGCCCGGAGGCTCCTGACCGAGCT 328
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 DB 329 TCCAAACCCGCTGCTGTCGACGAGCCCAAAATCCCATCCCGACAGTGTGCACTCAAG 388
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 DB 389 ACTAAGAAAGCACTGGCATCACTTACCTGGGACCTTCCTCGTGGGAGCTGGCGTG 448
 QY 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 DB 449 GCGGCTGGCTACTCTGGAAGTTTCATGGCAGCAGTGTCTCAACTCTGGGATAGAGTGC 508
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 DB 509 GACTCTCAGGTACCTGCGCATCAACCCCTCACTGCTGTGATGGCGTGCACACTGCCCC 568
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPhelIleLeuGlnVal 160
 DB 569 GGCGGGGAGGAGAGAAATCGGTGTGTGCTGCTTACGGACCAAACTTCATCTTCAGGTG 628
 QY 161 TyrSerSerGlnArgLysSerThrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
 DB 629 TACTCATCTCAGAGAGAGTCTGGCACCCTGTGGCCAAAGCAGCTGGAACGAGAACTAC 688
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPhetYrSerSerGlnGlyIle 200
 DB 689 GGGCGGCGGCTCGCAGGACATGGCTATAGAAATATTTTACTCTAGCAAGGAATA 748
 QY 201 ValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 DB 749 GTGGATGACAGGGGATCCCGACGCTTTATGAACTGAACACAACTGGCGGCAATGTCAT 808
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 DB 809 ATCTATAAAAACTGTACACAGTATGCTGTCTTCAAAAGCAGTGGTTCCTTTACGC 868
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 DB 869 TGTATAGCTTGGGGGTCACTTGAACTCAAGCCGCCAGCAGCAATTTGGCGGCGGAG 928

QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 DB 929 AGCGGCTCCCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 988
 QY 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
 DB 989 TCGGAGGCTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1048
 QY 301 ProLeuAsnAsnProTyrHisThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 DB 1049 CCTCTTACCAATCCATGGCATTTGGCGGCAATTTGGCGGCAATTTGGCGGCAATTT 1108
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 DB 1109 TTCTATGGAGCGGATACCAAGTAGAAAAAGTGAATTTCTCATCAAAATTTATGATCTCA 1168
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 DB 1169 ACCAAGACCAATGACATTTGGCTGATGAGCTGCAAGAGCTCTGACTTTCAACGACCTA 1228
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 DB 1229 GTGAAACCAAGTGTCTGCTGCCCAACCCAGGATGATGCTGCAGCCAGACAGCTCTGCTG 1288
 QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 DB 1289 ATTTCCGGGTGGGGGCGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGACGCTGCC 1348
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 DB 1349 AAGTGTCTTCTCATTTGAGACACAGATGACACAGCAGATATGCTCTATGACACCTGATC 1408
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 DB 1409 ACACGACGATGATCTGTGCGGCTTCTGTCAGGGGAACTGCTGATTTCTGCCAGGCTG 1468
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrIleIleGlyAspThrSer 460
 DB 1469 AGTGGAGGCGCTCTGGTCACTTCGAAAGAACATATCTGGTGGCTGATAGGGGATACAAGC 1528
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 DB 1529 TGGGGTCTGGCTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTATTTC 1588
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 DB 1589 ACGGACTGGATTTATCGACAAATGAGGCGCAGACGCGC 1624

RESULT 3

US-10-116-802-203
 ; Sequence 203, Application US/10116802
 ; Publication No. US20030065157A1

GENERAL INFORMATION:

; APPLICANT: Amy Lasek
 ; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
 ; FILE REFERENCE: PA-0045 US
 ; CURRENT APPLICATION NUMBER: US/10/116,802
 ; CURRENT FILING DATE: 2002-04-04
 ; PRIOR APPLICATION NUMBER: 60/281,593
 ; PRIOR FILING DATE: 2001-04-04
 ; NUMBER OF SEQ ID NOS: 519
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 203
 ; LENGTH: 2486
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No: 991163CB1
 US-10-116-802-203

Alignment Scores:

Pred. No.: 2,68e-306 Length: 2486
 Score: 2712.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.82% Indels: 0
 DB: 13 Gaps: 0

US-09-615-285B-2 (1-492) x US-10-116-802-203 (1-2486)

```

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 149 ATGGCTTTGAACACAGGGTACACACCTGCTATTGACCTTACTATGAAACCATGGATAC 208
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
DB 209 CAACCGGAAACCCCTATCCGACAGCCACTGTGGTCCCACTGTCTACGAGTGCAT 268
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 269 CCGGCTCAGTACTACCCGCTCCCGCTGGCCAGTACGCCCGGAGGCTCTGACGAGGCT 328
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 329 TCCAAACCCCGTCTGTGACGAGCCCAATCCCACTCCGAGCAGGTGTGCACTCAAG 388
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
DB 389 ACTAAGAAAGCACTGTGGCATCACCITGACCTTGGGACCTTCTCTGGGAGCTGCGCTG 448
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 449 GCGCTGGGCTACTCTGGAAGTTCTATGGGAGCAAGTCTCCAACTCTGGGATAGAGTGC 508
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 509 GACTCTCAGGTACCTGCAATCAACCCCTTAACTGGTGTGATGGGCTGCTACACTGCCCC 568
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 569 GCGGGGAGGAGCAGAGATCGGTGTCTGCTCCTACGGACCAAACTTCACTCTCAGGTG 628
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB 629 TACTCATCTCAGAGAGAGTCTTGGCACCTGTGTGCCAAGCAGCTGGAGCAGAGACTAC 688
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 689 GGGCGGGCGGCTCGAGGAGACATGGGCTATAAGATATATTTTACTCTAGCCCAAGGAATA 748
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 749 GTGGATGACAGCGGATCCACCACTTATGAACTGACACAACTGCGGCAATGCTGAT 808
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 809 ATCTATAAAAACTGTACCACAGTATGCTGTTCTTCAAAAGCAGTGGTCTTTTACGC 868
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 869 TGTATAGCTGTGGGGTCACTTGAACCTCAGCCGCCAGACAGAGTGTGGGCGCGAG 928
QY 261 SerAlaLeuProGlyAlaTyrProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
DB 929 AGCGGCTCCCGGGGCTGGGCTTGGCAGGTCTAGCTGACGCTGCCAGAGACGTCCACGCTG 988
QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
DB 989 TGGGAGGCTCATCATCATCCCCCGAGTGGCCGTGACAGCGGCCCACTGGTGGAAAAA 1048
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 1049 CCTCTTAAACATCCATGGCATGGACGGCATTTGGGGGATTTTGGAGCAATCTTTCTATG 1108

```

RESULT 4

```

US-10-205-823-414
; Sequence 414, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endegge, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: NRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 414
; LENGTH: 3226
; TYPE: DNA

```

ORGANISM: Homo sapiens
US-10-205-823-414

Alignment Scores:

Pred. No.: 5,2e-306 Length: 3226
Score: 2711.00 Matches: 490
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 15 Gaps: 0

US-09-615-285B-2 (1-492) x US-10-205-823-414 (1-3226)

Qy 1 MetAlaLeuAnSerGlySerProAlaLeuGlyProTyrTyrGluAsnHisGlyTyr 20
Db 128 ATGGCTTGAACATCAGGTCACCAACAGCTATTTGGACCTTATATGAAAACCATGGATAC 187
Qy 21 GluProGluAnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 188 CAACCGGAAACCCCTATCCCGACAGCCCACTGTGTGTCCTCCCACTGTCTACGAGTGCAT 247
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 248 CCGGCTCAGTACACCGTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 307
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 308 TCCAAACCCCGCTCTGTGACGACGCCCAATCCCAATCCCAATCCCAATCCCAATCCCAAT 367
Qy 81 ThrLysAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 100
Db 368 ACTAAGAAACATCTGTGACCTTGTACCTGTGGGACCTTCTCTGTGGGAGCTGGCTG 427
Qy 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyLysGluCys 120
Db 428 GCCGCTGGCTACTCTCGAAGATTCATGGGACGACGACGACGACGACGACGACGACGAC 487
Qy 121 AspSerSerGlyThrCysLysLeuAnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 488 GACTCTCTCAGTACCTGTCATCAACCCCTCTAATCTGTGTGATGGGCTGTACACTGCCCC 547
Qy 141 GlyGlyLeuAspGluAnArgCysValArgLeuTyrGlyProAsnPheLeuLeuGlnVal 160
Db 548 GGGGGGGAGACGAGAATCGGTGTGTGCTCTACGACCAAACTTCTCATCTCTCAGATG 607
Qy 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAnTyr 180
Db 608 TACTCATCTCAGAGGAAGCTCTGGACCTGTGTGCCAAGACGACTGGACGAGAACTAC 667
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
Db 668 GGGCGGGCGGCTGTCAAGGACATGGGCTATAAGATAATTTTTTACTCTAGCAAGGAATA 727
Qy 201 ValAspAspSerGlySerThrPheMetLysLeuAnThrSerAlaGlyAsnValAsp 220
Db 728 GTGGATGACAGCGGATCCACCACTTATGAACACTGAACACAGTCCCGGCAATGTCTGAT 787
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 788 ATCTATAAAAACTGTACCACTGATGCTGTCTTCAAAAGCAGTGTGTTCTTTTACGC 847
Qy 241 CysIleAlaCysGlyValAsnLeuAnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 848 TGTATAGCTTGGGGGTCACTTGAACCTCAAGCCCGACAGCAGATTTGTGGGCGGTAG 907
Qy 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
Db 908 AGGCGGCTCCCGGGGCGCTGGCCCTGGCAGGTTCAGCTTCAGCTCCAGAACTCCAGGTG 967
Qy 281 CysGlyGlySerIleLeuProGluTyrPheValThrAlaAlaHisCysValGluLys 300
Db 968 TGGGAGGCTCCATCATCAACCCCGAGTGGATCGTGACAGCGCGCCCACTGGGTGGAATA 1027

Qy 301 ProLeuAnAsnProTyrHisTyrThrAlaPheAlaGlyLeuArgGlnSerPheMet 320
Db 1028 CCTCTTAACAATCCATGTCATGGCGCATTTGGCGGATTTTGAGACAATCTTTTCATG 1087
Qy 321 PheTyrGlyValGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1088 TTCTATGGAGCGGATACCAAGTAGAAAAGTATTTCTCATCCAAATATGACTCCAG 1147
Qy 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1148 ACCAAGAACATGACATTCGCTGATGAAGCTGCAAGAGCCCTCTGACTTTCAACGACCTA 1207
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1208 GTGAACACAGTGTGTGCGCCACCCAGGATGATGCTGCAGCCAGAACAGCTCTGCTGG 1267
Qy 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1268 ATTTCGGGTGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAAGTGTGAAAGCTGCC 1327
Qy 401 LysValLeuLeuLeuLeuThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1328 AAGTGCTTCTCATTTGAGACACAGATGCAACACAGATATGCTCTATGACAACTGATC 1387
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1388 ACACAGCCATGATCTGTGCCGCTTCTGTGACAGGGAACGTCGATTTCTTGGCAGGCTGAC 1447
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
Db 1448 AGTGGAGGCGCTCTGTGTCTCTCGAAGAACAAATATCTGTGGTGGTATAGGGATACAGC 1507
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1508 TGGGGTTCTGTGCTGTGCCAAGCTTACAGACAGAGGTGTACGGAGTGTGATGTGATTC 1567
Qy 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
Db 1568 ACGACTGTGATTTATCGACAAATGAGGGCAGACGSC 1603

RESULT 5

US-09-759-143-931
; Sequence 931, Application US/09759143
; Patent No. US20020022248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stoik, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-759-143-931

Alignment Scores:

Pred. No.: 2,736-306 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-759-143-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTTGAACCTCAGGGTCCACACAGCTATTTGACCTTACTATGAAACCATGGATC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValProThrValHis 40
DB 61 CAACCGGAAACCCCTATCCGACAGCCCAATCCCTATCCGAGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACATACCCGCTCCCGCTGCCAGTACGCCCGGAGGGTCTTACACCGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCAACCCCGCTGCTGTCAGCGAGCCCAATCCCTATCCGAGCAGTGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAAGAAAGCACTGTGCATCCTTGCCTTGCCTGGGAGCTTCTCGTGGAGCTGCGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 301 GCGCTGCGCTACTCTGGAAGTTCTAGGCGAGCAGTGTCTCAACTCTGGATAGATGTC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 361 GACTCCTCAGGACCTGTGCATCAACCCCTCAACTGGTGTGATGGCTGTGCATCTGCC 420
QY 141 GlyGlyLysGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 421 GCGGGGAGGAGGAGATCGGTGTGTCTGCTCTACGATCAAACTTCACTCTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGAAAGTCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 541 GCGCGCGCGCTGTCAGGAGCATGGCTATAAGATAATTTTACTCTAGCCAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACAGCGGATCCACACAGCTTATGAACCTGAACACAAAGTCCGCGCAATGTCAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAACTGTACACAGTATGCTGTCTTCARAAAGCAGTGGTTCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTTGGGGTCAACTTGAACCTCAAGCCCGCAGACGAGATTTGGGGCGCGAG 780
QY 261 SerAlaLeuProGlyValAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 AGCGCGCTCCCGGGGCTGGCTGGCAGGTGTCAGCTGCCAGTCCAGAACGTCACGTCG 840
QY 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
DB 841 TCGGAGGCTTCATCATCCCCCGAGTGTGTCAGCGCCGCCCTCGTGTGAAAAA 900
QY 301 ProLeuAsnAsnProTyrTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

DB 901 CCTTTAACAATCCATGCGCATTTGCGGGGATTTGAGACAATCTTTCAATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGlyLysValLysSerHisProAsnTyrAspSerLys 340
DB 961 TTTCTATGAGCGGATACCAAGTAGAAAGTATTTCTCATCCAAATATGACTCCAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
DB 1021 ACCAAGAACAATGACATTTGCGCTGATGAAGCTGCAGAAAGCTCTGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGAACACAGTGTCTGTCCCAACCCAGGATGATGTGCGCCAGAACACGCTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluGlyLysThrSerGluValLeuAsnAlaAla 400
DB 1141 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTGTCTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
DB 1201 AAGGTGCTTCTCATTTGAGACACAGATGCAACAGCAGATATGCTATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
DB 1261 ACACAGCAGCATGATCTGTGCGGCTTCTGTCAGGGGAACGTCGATTTCTTGCAGGGTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnIleTrpTrpLeuIleGlyAspThrSer 460
DB 1321 AGTGAGGGCTCTGTGTCATCTCGAAGAACAATATCTGTTGGCTGATAGGGGATACAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
DB 1381 TGGGGTCTGCTGTGTCACAAAGCTTACAGACAGAGGTACGCGAATGTGATGGTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGGATTTATTCGACAAATAGGGCGACAGCGC 1476

RESULT 6

US-09-780-669-931
; Sequence 931, Application US/09780669
; Patent No. US2002005197A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darriack
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780.669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476


```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-780-669-931

Alignment Scores:
  2.73e-306      Length: 1476
Pred. No.:      2709.00      Matches: 491
Score:          99.80%      Conservative: 0
Percent Similarity: 99.80%      Mismatches: 1
Best Local Similarity: 99.71%      Indels: 0
Query Match:    9          Gaps: 0
DB:

US-09-615-285B-2 (1-492) x US-09-780-669-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAACCTCAGGCTCACCACGCTATTGGACCTTACTATGAATAACCATGGATAC 60

QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValProThrValHis 40
DB 61 CAACCGGAACCCCTATCCGACAGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 120

QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCCGTCGCCCGTCAGTACGCCCGGAGGGTCTCAGCAGGCT 180

QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 161 TCCAAACCCGCTGCTGCAGCAGGCCCAATCCCAATCCGAGCAGGTGTCACCTCAAG 240

QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAAGAAGACACTGTGCATCACCTTGACCTGGCGGACCTTCTCGTGGGAGCTGCCTG 300

QY 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 301 GCGGCTGGCGTACTCTGGAAGTTCATGGCAGCAAGTCTCAACTCTGGGATAGATGC 360

QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 361 GACTCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGGCTGCACACTGCCCC 420

QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleGlnVal 160
DB 421 GCGCGGAGGAGCAGAAATCGGTGTCTCGCCTCTACGGATCAAACTTCTCCTTCAGGTG 480

QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGGAAGTCTCTGGCACCCCTGTGTCAGACAGCACTGGACGAGACTAC 540

QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyTle 200
DB 541 GCGCGGCGGCTGCAGGACATGGCTATAAGAATAATTTTACTCTAGCCAAAGGAATA 600

QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACAGGGATCCACCAAGCTTTATGAACCTGAACACCAAGTCCCGGCAATCTCAT 660

QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAAACCTGTACACAGTGTATGCTGTCTTCAAAAGCAGTGTGTTCTTTACGC 720

QY 241 CysIleAlaCysGlyValAsnLeuAsnSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTTGGGGGTCAACTGAATCAAGCCGCGCAGAGGATGTGGCGCGCAG 780

QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 AGCCGCTCTCCGGGGGCTTGGCCCTGGCAGGTTCAGCTGCACGTCCAGAACCTCCACGTG 840

QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
DB 841 TGGGAGGCTCCATCATCAACCCCGAGTGGATCGTGACAGCCGCCCACTCGCTGGTGAATA 900
```

```
301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
901 CCTCTTAAACATCCATGGCATTTGGACGGCATTTGGGGGATTTTGAGACAACTTTCATG 960

321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
961 TTTCTATGGAGCGGATACCAAGTAGAAAAAGTGAATTTCTCATCCAAATATGACTCCAAG 1020

341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
1021 ACCAGAACATATGATTCGGCTGTGAAGCTGCAGAGCCTTGACTTTCAACGACCTA 1080

361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
1081 GTGAACACCACTGTCTGCCCAACCCAGGCATGATGCTGCAGCCAGACAGCTCTGCTGG 1140

381 IleSerGlyTyrGlyAlaThrGluGlyLysGlyLysThrSerGluValLeuAsnAlaAla 400
1141 ATTTCGGGTGGGGGCCACCCAGGAGAAAGGAGACCTCAGAGTGTGTGAACGCTGCC 1200

401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
1201 AAGTGTCTTCTCATTTGAGACACAGAGATGCCAACAGCAGATATGCTATGACAACTGATC 1260

421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
1261 ACACCAAGCCATGATCTGTCCGGCTTCTTCGAGGGGAGACGTGCAATTTCTGCCAGGTGAC 1320

441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
1321 AGTGGAGGCTCTGTGTCATTCGAGAGAACATATCTGGTGGCTGATAGGGATACAAGC 1380

461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
1381 TGGGATTTCTGGCTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGGTATTC 1440

481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
1441 ACGACTGGATTTATCAACAAATGAGGGCAGACGGC 1476

RESULT 7
US-09-822-827-931
; Sequence 931, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-827-931

Alignment Scores:
Pred. No.:      2.73e-306      Length: 1476
Score:          2709.00      Matches: 491
Percent Similarity: 99.80%      Conservative: 0
Best Local Similarity: 99.80%      Mismatches: 1
Query Match:    99.71%      Indels: 0
DB:              9          Gaps: 0

US-09-615-285B-2 (1-492) x US-09-822-827-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAACCTCAGGCTCACCACGCTATTGGACCTTACTATGAATAACCATGGATAC 60
```


QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValProThrValHis 40
DB 61 CAACGGAAACCCCTATCCCGACAGCCCACTGGGTCCTGCTACGAGTGCAT 120
QY 41 ProAlaGlnTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCGTCCCGTCCCGCAGTACGCCCCGAGGGTCTTGACGAGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCAACCCCGTCTGTCAGCAGCCCAATCCCATCCGGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysLeuAlaLeuCysLeuThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
DB 241 ACTAAGAAAGCACTGTGCATCACTTGCCTGGGACCTTCTCTGCGGAGCTGGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyLeuGluCys 120
DB 301 GCGCTGGCCTACTCTGGAGTTTCATGGGAGCAAGTGTCTCAACTCTGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 361 GACTCTCAGGTACCTGCATCAACCCCTCAACTGCTGATGGCGTGTCACTGCCCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheLeuGlnVal 160
DB 421 GCGGGGAGGAGCAAGTCTGGTGTCTGCTCTACGATCAAACTCTCACTCTCAGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGAGTCTGGCACTCTGTGTGCCAAGACGACTGGACGAGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
DB 541 GCGGGGCGGCTGTCAGGAGCACTGGCTATAGAAATTAATTTTACTCTAGCCAGGAATA 600
QY 201 ValAspSerSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACAGCGGATCCACGACTTTATGAACCTGAACCAAGTGGCGGCAATGTCTG 660
QY 221 IleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAACTCTACACAGTATGCTGCTCTTCAAAAGCAGTGGTTCCTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
DB 721 TGTATAGCTCGCGGCTCAACTTTGAACCTCAAGCCGACAGCAGGATTTGTGGCGGCGAG 780
QY 261 SerAlaLeuProGlyAlaTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 AGCGGCTCCCGGGGCTGGCCCTGGCAGTTCAGCTGACGTCACGACGACGTCACGCTG 840
QY 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
DB 841 TCGGAGGCTCCATCATCCCCCGAGTGGATCGTGACAGCGGCCCACTGCTGGGAAAA 900
QY 301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CTCTTAACTATCCATGGCATGGACGCAATTTGGGGATTTTGACAACTCTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
DB 961 TTTCTATGGAGCGGATCAAGTAGAAAAAGTATTTCTCATCCAAATTTAGTCTCAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
DB 1021 ACCAAGAAACAATGATTCGCTGATGAGCTGGCAGAGCCCTGATCTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGAACCACTGTGTCTGCCCAACCCAGGATGATGCTGACGACGACGACGCTCTGCTG 1140

QY 391 IleSerGlyTyrGlyAlaThrGluGluLysGlySerThrSerGluValLeuAsnAlaAla 400
DB 1141 ATTCCGGGTGGGGGCCACCGAGAGAAAGGAGACCTCAGAAGTGTGACACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuLe 420
DB 1201 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGluGlyAsnValAspSerCysGlnGlyAsp 440
DB 1261 ACACAGGCACTGATCTGTGGCGTCTCTGAGGGAACTGATTTCTCCAGGGTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
DB 1321 AGTGAGGCTCTGCTGCTCACTTCGAGAAACAATATCTGTGGTGTATAGGGGATACAAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
DB 1381 TGGGGTCTGCTGTGCCAAGCTTACAGACAGGAGTGTACGGGAATGTGATGGTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGGATTTATCGACAAATGAGGGCGACGCGC 1476

RESULT 8

US-09-895-793-931
Sequence 931, Application US/09895793
Publication No. US20020192763A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuciu
APPLICANT: Kalos, Michael D.
APPLICANT: Ketter, Marc W.
APPLICANT: Stoik, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedwick, Thomas S.
APPLICANT: Carter, Derrick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Repler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.534C2
CURRENT APPLICATION NUMBER: US/09/895,793
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 931
TYPE: DNA
ORGANISM: Homo sapiens
US-09-895-793-931

Alignment Scores:
Pred. No.: 2,73e-306 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservatives: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-895-793-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 1 ATGGCTTTTGAACCTCAGGTCACACAGCTATTGGACCTTACTATGAAACCATGZATC 60
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 Db 61 CAAACCGAAACCCCTATCCGACACAGCCCACTGTGTCCCACTGTCTACGAGGTGCAT 120
 QY 41 ProAlaGlyTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CCGGCTCAGTACTACCCGTCGCCGTCGCCAGTACGCCGAGGCTCCTGACGAGGCT 180
 QY 61 SerAsnProValValCysThrGlnProTyrSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCCAAACCCGTCGTCTGACGACGAGCCCAATCCCAATCCGACGAGTGTGACCTCAAG 240
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 241 ACTAAGAAAGCACTGTGTCATCACTTGCACCTGGGGAGCTTCTCCTGGTGGAGCTGCGCTG 300
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 301 GCGCTGGCTACTCTGGAAGTTCTGGGACAGAGTCTCCAACTCTGGGATAGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 361 GACTCTCAGGTACCTGTCATCAACCCCTTAACTGTGTGATGGCGTGTCACTGCCCC 420
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 421 GCGGGGAGCAGAGATCGGTGTGTTCGCTCTACGGATCAAACTTCATCTTCAGGTG 480
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 Db 481 TACTCATCTCAGAGGAAGTCTGGGACCCCTGTGTGCCAAGACGACTGGAACGAGAACTAC 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
 Db 541 GCGGGGCGGCTGTGAGGACATGGCTATAGATATATTTTACTCTAGCCAGGATTA 600
 QY 201 ValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAAACTGGAACACAGTGGCGGCAATGTGCAT 660
 QY 221 IleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 661 ATCTATAAAACTGTATACACAGTATGCTGTCTTCAAAGCAGTGGTTCCTTACGC 720
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 721 TGTATAGCTGGGGGTCAACTTGAATCAAGCGCCAGAGCAGGATTTGTGGCGCGAG 780
 QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 781 AGCGGCTCCCGGGGCTGGCCCTGGCAGGTGACGCTGCGACGTCGAGACGTCACGTCG 840
 QY 281 CysGlyGlySerIleThrProGluThrPheValThrAlaHisCysValGluLys 300
 Db 841 TCGGAGGCTCCATCATCCCGCGAGTGGATCGTGACAGCGCGCCCACTGCGTGGAAAA 900
 QY 301 ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 901 CCTCTTAACATCCATGGCATTTGGACGGCATTTGGGGGATTTTGAGCAATCTTTCATG 960
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 961 TTCTATGGAGCGGATACCAAGTAGAAAAAGTGAATTTCTCATCCAAATTTATGACTCCAA 1020
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1021 ACCAAGAACATGACATTTGGCTGATGAAGCTGAGAGGCTCTGACTTTTCAACGACCTTA 1080

QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 Db 1081 GTGAACACCATGTGTCTGCCCAACCCAGGATGATGTGTGCAGCCAGAACAGCTCTGCTGG 1140
 QY 381 IleSerGlyTyrGlyValThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 Db 1141 ATTTCCGGGTGGGGGCGCCACGAGGAGAAAGGAGACCTCAGAGTGTCTGAAAGCTGCC 1200
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1201 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACACACCTGATC 1260
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1261 ACACGACCATGATCTGTCCGGCTTCTGTGCGGGACGTCGATTTCTTGGCCAGGGTGAC 1320
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
 Db 1321 ACTGGAGGCGCTCTGTCTCACTTCGAAGAACAATATCTGGTGGCTGATAGGGGATACAAGC 1380
 QY 461 TrpGlySerGlyCysAlaAlaValAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1381 TGGGTTCTGGCTGTGCCAAAGCTTACAGACAGAGAGTGTACGGGAATGTGATGTTATTC 1440
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 Db 1441 ACCGACTGGATTTATCGCAAAATGAGGCGACAGCGC 1476

RESULT 9

US-09-895-814-931
 ; Sequence 931, Application US/09895814
 ; Publication No. US20020193296A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, David C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yudi
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedwick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Hural, John
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Vinals de Bassola, Carlota
 ; APPLICANT: Foy, Teresa
 ; APPLICANT: Fanger, Gary R.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 ; FILE REFERENCE: 210121.427C26
 ; CURRENT APPLICATION NUMBER: US/09/895,814
 ; CURRENT FILING DATE: 2001-06-29
 ; NUMBER OF SEQ ID NOS: 990
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 931
 ; LENGTH: 1476
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-895-814-931

Alignment Scores:
 Pred. No.: 2,73e-306 Length: 1476
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1

Query Match: 99.71% Indels: 0
DB: 9 Gaps: 0
US-09-615-285b-2 (1-492) x US-09-895-814-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAATCACTCAGGGTCCACCCAGCTTACTATGAAACCACTGATGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValProThrValHis 40
DB 61 CAACCGGAAACCCCTATCCCGACAGCCCACTGGTCCCACTGCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CGGGCTCAGTACTACCCGCTCCCGCTGCCAGTACGCCCGGGGCTCCGACGCGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCCAAACCCGCTGCTGTACAGCGAGCCCAATCCCCATCCGGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAAGAAAGCACTGTGCATCACTTACCTTGGGAGCTTCTTCTGGGAGCTGGCGTG 300
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 301 GCGCTGGCTACTCTGGAAGTTCTATGGGAGCAAGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
DB 361 GACTCTCAGTACTGTGATCAACCCCTTACTTGGTGTGTGATGGCTGTGCATCTGCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 421 GCGGGGAGGAGCAGATCGGTGTCTGCTTACCGATCAAACTTCTATCTTCTTACGGTG 480
QY 161 TyrSerSerGlnArgLysSerTroHisProValCysGlnAspTyrAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGAAAGTCTGGGACCCCTGTGTGCCAAGCACTGGAGACGAGACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 541 GGGCGGGCGGCTGTGAGGAGACATGGGCTATAGAAATATTTTACTCTAGCCCAAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACAGCGGATCCACCACTTTATGAACTGACCAAGTGGCGGCAATGTGAT 660
QY 221 IleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAAACTGTACACAGTATGCTTCTTCAAAAGCAGTGGTTCCTTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
DB 721 TGTATAGCTGGGGGTCACTTGTACTCAGCGCCAGAGCAGGATGTGGCGGCGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 AGCGGCTCCCGGGGCGCTGGCGCTGGAGGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 840
QY 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
DB 841 TCGGAGGCTCATCATCACTCCCGAGTGGATCTGTGAGCGCGGCTCGTGGTGGAAAAA 900
QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CTTCTTAAACAATCCATGAGTGGACGGCATTTGCGGGATTTTGACAACTTTTCTATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
DB 961 TTCTATGAGCGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATTTATGACTCAAG 1020

QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
DB 1021 ACCAAGAACATGACATTGGCTGTATGAAGCTGCAGAGGCTCTGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
DB 1081 GTGAACCACTGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGAAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
DB 1141 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTGTGTAACCTGCC 1200
QY 401 LysValLeuLeuLeuGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuLe 420
DB 1201 AAGTGTCTTCAATTGAGACACAGATGCAACAGCAGATATGTCTATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
DB 1261 ACACGACCATGATCTGTGCGGCTTCTTGCAGGGGACGTCGATTTCTTGCAGGGTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
DB 1321 AGTGGAGGGCTCTGTGCTACTTGCAGAAACAATATCTGTGTGTATAGGGGATCAAGC 1380
QY 461 TrrGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
DB 1381 TGGGTCTCTGGCTGTGCCAAAGCTTACAGCAGGAGTGTACGGGAATGTGTATGTGTTTC 1440
QY 481 ThrAsnTrrPileTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGACTGGATTTATCGACAAATGAGGGGAGACGGC 1476

RESULT 10

US-10-012-896-931
; Sequence 931, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012, 896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-896-931

Alignment Scores:

Pred. No.: 2,73e-306 Length: 1476
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.71% Indels: 0
 DB: 14 Gaps: 0

US-09-615-285B-2 (1-492) x US-10-012-896-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 DB 1 ATGGCTTTGAACTCAGGGTCAACCAAGCTATGAGCTTACTATGAAMACCAGGATAC 60
 QY 21 GluProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 DB 61 CAACCGGAAACCCCTATCCGACAGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 DB 121 CCGGCTCAGTACTACCCGCTCCCGCTGCCAGTAGTACGCCGAGGGTCTTGACGAGGCT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 DB 181 TCCAAACCCGCTGCTGCGACGAGCCCAATCCCAATCCCACTCCGGACAGTGTGCACCTCAAG 240
 QY 81 ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyValAlaLeu 100
 DB 241 ACTAAGAAAGCAGTGTGCATCACTTACCTGGGGGACCTTCTCGTGGGAGCTGCCGTG 300
 QY 101 AlaAlaGlyLeuLeuTyrPhePheMetGlySerLysCysSerAsnSerGlyIleGlyCys 120
 DB 301 GCCGCTGGGCTACTCTGGGAAGTTCATGGGACAGTGTCTCAACTCTGGGATAGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 DB 361 GACTCTCAGTACCTGTGCATCAACCCCTCAACTGCTGTGTGTGTGTGTGTGTGTGTGTGT 420
 QY 141 GlyGlyGluAsnGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 DB 421 GCGGGGAGACAGAGATCGGT 480
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
 DB 481 TACTCATCTCAGAGGAAGTCTCTGGCACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 DB 541 GCGCGGCGGCTGCGAGGACATGGCTATAGAAATATTTTACTCTAGCCAGGAGATA 600
 QY 201 ValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 DB 601 GTGGATGACAGCGGATCCACCAAGCTTTATGAACCTGAAACACAGTGTGGCGCAATGTGAT 660
 QY 221 IleTyrLysValLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 DB 661 ATCTATAAAACTGTACCACAGTATGCTGTCTTCAAAGCAGTGGTGTCTTTTACGC 720
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 DB 721 TGTATAGCTGCGGGGTCAACTGAACCTCAAGCGCGCACAGCAGGATGTGTGGCGCGAG 780
 QY 261 SerAlaLeuProGlyValAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 DB 781 AGCGGCTCCCGGGGCTTGGCCCTGGCAGGTGACCTTGCACGTCCAGAACGTCACGCTG 840
 QY 281 CysGlyGlySerIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
 DB 841 TCGGAGGCTCCATCATCAACCCCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
 QY 301 ProLeuAsnAspProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

DB 901 CCTCTTAAACAATCCATGGCATTTGGACGGCATTTTGGGGGATTTTGAGACAATCTTTTCATG 960
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 DB 961 TTCTATGGAGCGGATCCCAAGTAGAAAAAGTATTTCTCATCAATATATGACTTCAAG 1020
 QY 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 DB 1021 ACCAAGAACATGACATTTGGCTGTATGAAGCTGCAGAAAGCCTCTGACTTTTCAACGACCTA 1080
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 DB 1081 GTGAACACCGTGTGTGTGCCCAACCCAGGATGTGTCTGCAGCCAGAACAGCTTCTGCTGG 1140
 QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 DB 1141 ATTTCCGGGTGGGGGGCCACCCGAGGAGAAAGGAAGACCTCAGAAGTGTGTGAACGCTGCC 1200
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 DB 1201 AAGGTGCTTCTCATTTGAGACACAGAGATGCCAAGCAGATATGTCTATGACACCTGATC 1260
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 DB 1261 ACACGAGCGCTCTGT 1320
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
 DB 1321 AGTGGAGGGCTCTGT 1380
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 DB 1381 TGGGGTCTGT 1440
 QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
 DB 1441 ACGGACTGGATTTATTCGACAAATGAGGCGACAGCGGC 1476

RESULT 11

US-10-144-678A-931
 ; Sequence 931, Application US/10144678A
 ; Publication No. US20030157089A1

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
 APPLICANT: Dillon, Davin C.
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Harlocker, Susan L.
 APPLICANT: Jiang, Yuqiu
 APPLICANT: Henderson, Robert A.
 APPLICANT: Kalos, Michael D.
 APPLICANT: Fanger, Gary R.
 APPLICANT: Retter, Marc W.
 APPLICANT: Stolk, John A.
 APPLICANT: Day, Craig H.
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Carter, Darrick
 APPLICANT: Li, Samuel X.
 APPLICANT: Wang, Aljun
 APPLICANT: Skeiky, Yasar A. W.
 APPLICANT: Hepler, William T.
 APPLICANT: Hural, John
 APPLICANT: McNeill, Patricia D.
 APPLICANT: Houghton, Raymond L.
 APPLICANT: Vinals y de Baesols, Carlota
 APPLICANT: Foy, Teresa M.
 APPLICANT: Watanabe, Yoshinhiro
 APPLICANT: Deng, Ta
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 ; FILE REFERENCE: 210121.427C28
 ; CURRENT APPLICATION NUMBER: US/10/144,678A
 ; CURRENT FILING DATE: 2002-08-12
 ; NUMBER OF SEQ ID NOS: 1033

FASTSEQ for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-144-678A-931

Alignment Scores:
Pred. No.: 2,73e-306 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 15 Gaps: 0

US-09-615-285B-2 (1-492) x US-10-144-678A-931 (1-1476)

```
QY 1 MetAlaLeuAenSerGlySerProAlaileGlyProTyrTyrGluAenHisGlyTyr 20
DB 1 ATGGCTTTGAACCTCAGGGTACCACACAGCTATTGGACCTTACTATGAACACCTGATAC 60
QY 21 GlnProGluAenProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 61 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCCGTCCTCCCGTGGCCAGTACGCCCGCGAGGGTCTTGACGCGAGGCT 180
QY 61 SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCAACCCCGCTGCTGACGAGGAGCCCAATCCCATCCGAGGAGGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysileThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
DB 241 ACTAAGAAAGACTGTGATCACCTTGACCTCGGGACCTTCTCGTGGGAGCTGGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAenSerGlyLeuGlyCys 120
DB 301 GCGCTGGCCCTACTCTGAAGTTATGGGAGCAGGTGCTCCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysileAenProSerAenTrpCysAspGlyValSerHisCysPro 140
DB 361 GACTCCTCAGGTACCTGATCAACCCCTCTAAGTGGTGTGATGGCGTGTCACTGCCCC 420
QY 141 GlyGlyGluAenGluAenArgCysValArgLeuTyrGlyProAenPheLeuGlnVal 160
DB 421 GCGGGAGGAGGAGGAGATCGGTGTGTCCTCTACGATCAAACTTCATCCTTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspTrpAenGluAenTyr 180
DB 481 TACTCATCTCAGAGGAGTCTGTCGACCCCTGTGTCACAGCAGTGGAACTGAGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAenAenPheTyrSerSerGlnGlyLe 200
DB 541 GGGCGGGCGGCTCAGGAGCATGGGCTATAGAAATATATTTTACTCTAGCCAGGAATA 600
QY 201 ValAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAsp 220
DB 601 GTGGATGACAGCGGATCCACAGCTTTATGAAATGAACAAAGTGGCGGCAATGTGCAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAGAACTGTACACAGTGTGCTGTCTTCTCAAAAGCAGTGGTTCCTTACGC 720
QY 241 CysIleAlaCysGlyValAenLysAenSerSerArgGlnSerArgLeValGlyGlyGlu 260
DB 721 TGTATAGCTGGGGGTCAACTTGAACCTCAAGCGCCAGAGGAGTGTGGGCGCGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAenValHisVal 280
DB 781 AGCGGCTCCCGGGGGCTGACCTGGCAGGTGACGCTGACGCTCCAGAGACGCTCCACGCTG 840
```

```
QY 281 CysGlyGlySerIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
DB 841 TCGGAGGCTCCATCATCATCCCCGAGTGGATCGATCGTACAGCCGCCCATCTGCGTGAATAA 900
QY 301 ProLeuAenAenProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CCTCTTAACAATCCATGGCATTTGGACGGCATTTGGGGGATTTTGAGACAAATCTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGlyLysValIleSerHisProAenTyrAspSerLys 340
DB 961 TTCTATGAGCGGATACCAAGTAGAAAAGATTTCTCATCCAAATATGACTCCCAAG 1020
QY 341 ThrLysAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu 360
DB 1021 ACCAAGAACCAATGACATTCGCTGATGAAGCTGCAAGAGCCTCTGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGAACCAAGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAenAlaAla 400
DB 1141 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAAGTGTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrValTyrAspAenLeuIle 420
DB 1201 AAGGTGCTTCTCATTCAGACACAGAGATGCAACAGAGATATGTCTATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAspSerCysGlnGlyAsp 440
DB 1261 ACACGAGCATGATCTGTCCGCTTCTGCGAGGAGAACTGATCTTTCAGGAGTGCAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAenAenIleTyrTrpLeuIleGlyAspThrSer 460
DB 1321 AGTGGAGGGCTCTGGTCACTTCGAAGAACATAATCTGGTGGCTGATAGGGGATACAAGC 1380
QY 461 TrpGlySerCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe 480
DB 1381 TGGGTTCTGGCTGTGCCAAAGCTTACACACAGAGGTACCGGAATGTGATGGTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGGATTTATCGACAAATGAGGGCAGACGGC 1476
```

RESULT 12

US-10-294-025-931
; Sequence 931, Application US/10294025
; Publication No. US20030185830A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C29
; CURRENT APPLICATION NUMBER: US/10/294,025
; CURRENT FILING DATE: 2002-11-12
; NUMBER OF SEQ ID NOS: 1038
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 931
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-294-025-931

Alignment Scores:
Pred. No.: 2,73e-306 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 15 Gaps: 0

US-09-615-285B-2 (1-492) x US-10-294-025-931 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAATCAGCTCAGGCTCACCACAGCTATTGGACCTTACTATGAANAACCTGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 61 CAACCGGNAAGCCCTATATCCGACAGACGCCACTGTGGTCCGCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CGGCTCAGTACTACCGCTCCCGCTGCGCCAGTACGCCCGGAGGCTCTCAGCAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCCGCTGCTGCGACGAGCCCAATCCCATCCCGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyValAlaLeu 100
Db 241 ACTAAGAAAGCACTGTGCATCACTTGACCTGGGACCTTCTCTGTTGGAGCTGGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCGCTGGGCTACTCTCGAAGTTCATGGGACAGCAAGTGTCTCAACTCTCTGGGATAGATGC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGGTACCTGTGCATCAACCCCTCTAATCTGTGTGTGGGTGTCACTGTGCCCC 420
QY 141 GlyGlyAlaAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 421 GCGGGGAGGAGCAGAATCGGTGTGTGTGCTTACCGATCAAACTTCTCATCTTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGAGTCTCTGGACCCCTGTGTGCTCCAGACGACTGGACGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GCGCGGCGGCTCTGCGGACATGGCTATAGAAATAATTTTACTCTAGCAAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyValAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAGTGGCGGCAATGTGCAT 660
QY 221 IleTyrLysAlaLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAAACCTGTACCACAGTGTGCTGTCTTCAAAAGCAGTGGTTCCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
Db 721 TGTATAGCTCGGGGTCAACTTGAACCTCAAGCGCCAGACGAGCAATTTGTGGGCGGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGCGCTCCCGGGGCTCTGGCCCTGGGACGTCAGCTGCACGCTCCAGAACGTCACGCTG 840
QY 281 CysGlyGlySerIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
Db 841 TCGGAGGCTCCATCATCATCCCCCGAGTGGATCGTGACAGCGGCCACTGTGGTGAATAA 900
QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACAATCCATGGCATTTGGACGCGCATTTGGGGGATTTTGAGACATCTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTTCTATGAGCGCGGATACCAAGTAGAAAAGTATTTCTCATCCCAATATTAGCTCCAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGATTCGCTGTATGAAGCTGACAGAGCCCTCTGACTTTTCAACGACCTA 1080

QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1081 GTGAACCCAGTGTGTCTGCCAACCCAGGCATGATGTGCAGCCAGAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCGGGGTGGGGGCCACCGAGGAGAAAGGAAGACCTCAGAAAGTGTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGGTGCTTCTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGCAACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerGlnGlyAsp 440
Db 1261 ACACGAGCAGTATCTGTGCGGCTTCTGCGGGGACGTCGATTTTTCGACGGTGC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGCGCTCTGTCTCACTTCGAAAGAACAAATATCTGTGGCTGTAGGGGATACAAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGTTCTGCTGTGCTGCAAGCTTACAGACGAGGTGTACGGGAATGTGATGTGTTTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTATTCGACAAATGAGGCGAGACGGC 1476

RESULT 13

US-09-759-143-930
; Sequence 930, Application US/09759143
; Patent No. US2002002248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 930
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-759-143-930

Alignment Scores:
Pred. No.: 2,74e-306 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: Gaps: 0

US-09-615-285B-2 (1-492) x US-09-759-143-930 (1-1479)

QY 1 MetAlaLeuAenSerSerglySerProProAlaIleGlyProTyrTyrGluAenHisGlyTyr 20
DB 1 ATGGCTTTTGAATCAGGCTCACCACAGCTATTGGACCTTACTATGAAACACCATGATAC 60
QY 21 GlnProGluAenProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 61 CAACCGGAARACCCCTATCCCGCAGCAGCCACACTGTGTGCCCACTCTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCGCTCCCGCTCCCGCAGTACGCCCCGAGGCTCTGACGAGGCT 180
QY 61 SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCCAAACCCCGCTGCTGCAAGCAGCCCAATCCCAATCCGCGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysValAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAGAAAGACACTGTGCATCCTTGACCTCGGGGACCTTCTCTGTGGAGCTGGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerIysCysSerAenSerGlyIleGlyCys 120
DB 301 GCGCTGCGCTACTCTGAAATTCATGGCAGCAAGTGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerglyThrCysIleAenProSerAenTyrCysAspGlyValSerHisCysPro 140
DB 361 GACTCCTCAGTACTCTGATCAACCCCTCTAATCTGGTGTGATGGCTGTCACTGCCGCC 420
QY 141 GlyGlyLeuAspGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal 160
DB 421 GCGGGGAGGACGAGATCGGTGTGTTCGCTCTACGGATCAAACTTCATCTTCAGGTG 480
QY 161 TyrSerSerglnArgLysSerTyrPheProValCysGlnAspAspTrpAenGluAenTyr 180
DB 481 TACTCATCTCAGAGAAAGTCTGGCACCTGTGTGCCAGACACACTGGAAACGAGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrIlyAenAenPheTyrSerGlnGlyIle 200
DB 541 GCGCGGGCGGCTCGCAGGACATGGCTATAAGAAATAATTTTACTCTAGCCCAAGGAATA 600
QY 201 ValAspAspSerglySerThrSerPheMetLysLeuAenThrSerLacIlyAenValAsp 220
DB 601 GTGGATGACAGGATCACACAGCTTTTGAATCTGACACACAGTCCCGGCAATGTGCAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAAACTGTACCAAGTGTGCTGTCTTCAAAAGCAGTGTCTTTTACGC 720
QY 241 CysIleAlaCysGlyValAenAenSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTCGGGGGTCACTTGAACCTCAAGCGCCGACAGCAGGATTTGTGGCGCGCAG 780
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAenValHisVal 280
DB 781 AGCGGCTCCCGGGGGCTGGCTGGCAGGTGAGCTGACGCTCCAGAACGTCCACGCTG 840
QY 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGlyLys 300
DB 841 TGGGAGGCTCCATCATCACCCCGAGTGGATGTGTGACAGCGCCCACTGCGTGGAAAAA 900
QY 301 ProLeuAenAenProTyrHisThrThrAlaPheAlaGlyIleLeuArgLysPheMet 320
DB 901 CCTCTTAACAATCATGCGATGACGCGCATTTGCGGGGATTTTGAGACAATCTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGlyLysValIleSerHisProAenTyrAspSerLys 340
DB 961 TTCTATGGAGCGGATACCAAGTAGAAAAAGTGTATCTCATCAATTAATATGACTCCAAG 1020
QY 341 ThrLysAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu 360
DB 1021 ACCAAGAACAAATGATGCGCTGATGAAGCTGCAAGAGCTCTGTGACTTTTCAACGACCTA 1080

QY 361 ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
DB 1081 GTGAAACCAAGTGTGTCTGCCCAACCCAGCATGATGTCTGCAGCCAGAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValValLeuAenAlaAla 400
DB 1141 ATTTCCGGGTGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAGTGTCTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrValTyrAspAenLeuIle 420
DB 1201 AAGTGTCTTCTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGACACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAspSerCysGlnGlyAsp 440
DB 1261 ACACCAAGCATGATCTGTGCCGCTCTCTGCAGGAGAAAGCTCGATTTCTTGCAGGGTGC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAenAenIleTyrTrpLeuIleGlyAspThrSer 460
DB 1321 AGTGGAGGCGCTCTGGTCACTTCGAGAGAACATATCTGGTGGCTGATAGGGGATACAAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe 480
DB 1381 TGGGGTCTTGGGTGTGCCAAAGCTTTACAGACCAAGAGTGTACGGGAAATGTGTATGGTATTC 1440
QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGGATTTATCGACAAATGAGGGCAGACGGC 1476

RESULT 14

US-09-780-669-930
; Sequence 930, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Koughron, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 930
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-780-669-930

Alignment Scores:
Pred. No.: 2,746-306 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: Gaps: 0

US-09-615-285B-2 (1-492) x US-09-780-669-930 (1-1479)

```

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAACCTCAGGCTCACCACCGAGCTATTTGGACCTTACTATGAAACCATGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 61 CAACCGGAAACCCCTATCCCGACACAGCCCACTGTGTGCTCCCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCCGCTCCCGCTGCCAGTACCGCCGAGGCTCTCCGACGAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 181 TCCAAACCCCGCTGTCTGCAGCAGCCCAATCCCACTCCCGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAAGAAAGCACTGTGCATCACTTGAACCTTGGGAGCTTCTCCGTGGGAGCTGCGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 301 GCGCTGGCTACTCTGGAGTTCATGGCAGCAAGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnThrProCysAspGlyValSerHisCysPro 140
DB 361 GACTCCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGTGTGCACATGCCCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 421 GCGGGGAGAGAGATCGGTGTGTGCTTACGGATCAAACTTCATCTCCCTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGGAAGTCTGGCACTGTGTGCCAAGCAGCAGCTGGAAACGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
DB 541 GGGGGGGGGCTGAGGACATGGGCTATPAGATATATTTTACTCTAGCCAAAGGATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACAGCGGATCCACAGCTTTATGAACTGAAACACAAGTGGCGGCAATGTGGAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 AICTATAAAAACTGTACACAGTGTGCTTCTTCAAAAGCAGTGGTTCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTGCGGGGTCAACTTGAACCTCAAGCGCCACAGCAGGATTTGGCGGCGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 ACGGCGCTCCCGGGGCGCTGGCCCTGGCAGGTGACCTTCAGCTCCAGAACGTCACGTT 840
QY 281 CysGlyGlySerIleThrProGlnTyrPheValThrAlaAlaHisCysValGluLys 300
DB 841 TGGCGAGGCTCATCACCCTCCGAGTGGATCGTGACAGCGCGCCCACTCGCTGGTGGAAAA 900
QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CCTCTTAACATCCATGGCATTTGGAGCGCATTTGGGGGATTTTGGACACATCTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
DB 961 TTCTATGGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATTTATGATCCAAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360

```

```

DB 1021 ACCAAGAACATGACATTCGCTGATGAAGCTGCAGAGAGCTCTGACTTTTCAACGACCTA 1080
QY 361 VallysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
DB 1081 GTGAACACCATGTGTCTGCCAACCCAGGCAATGATGTGCAGCCAGAACAGCTCTGTCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
DB 1141 ATTTCCGGGTGGGGGCCCTCCGAGGAGAAAGGAGACCTCAGAGTGTCTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
DB 1201 AAGGTGCTTCTCATTCAGACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
DB 1261 ACACGAGCCATGATCTGTGCGGCTCTCTGCAGGGGAACGTCGATTTCTTCCAGGGTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
DB 1321 AGTGGAGGGCCTCTGGTCACCTTCGAAGAACAAATATCTGGTGGCTGTATAGGGGATACAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
DB 1381 TGGGGTCTGGCTGTGCCAACAGCTTACAGCACGAGGTGTACGGGAATGTGATGTTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGATTATTCGACAAATGAGGCGACAGCGC 1476

```

RESULT 15

US-09-822-827-930
 ; Sequence 930, Application US/09822827
 ; Patent No. US20020081680A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.534C1
 ; CURRENT APPLICATION NUMBER: US/09/822,827
 ; CURRENT FILING DATE: 2001-03-28
 ; NUMBER OF SEQ ID NOS: 982
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 930
 ; LENGTH: 1479
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-822-827-930

Alignment Scores:	2,74e-306	Length:	1479
Pred. No.:	2709.00	Matches:	491
Score:	99.80%	Conservative:	0
Percent Similarity:	99.80%	Mismatches:	1
Best Local Similarity:	99.71%	Indels:	0
Query Match:	9	Gaps:	0

US-09-615-285B-2 (1-492) x US-09-822-827-930 (1-1479)

```

QY 1 MetalAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAACCTCAGGCTCACCACCGAGCTATTTGGACCTTACTATGAAACCATGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 61 CAACCGGAAACCCCTATCCCGACACAGCCCACTGTGTGCTCCCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCCGCTCCCGCTGCCAGTACGCCCGAGGCTCTGACGAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

```

181 TCCAAACCCCGTCTGTCAGCAGGCCCAATCCCATCCGGGACAGTGTGCACCTCAAAG 240
QY ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db ACTAAGAAGACACTGTGCATCACTTGACCTGGGGACCTTCCTCGTGGAGCTGCGCTG 300
QY AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db GCGGTGGCTACTCTGGAAGTTTCATGGGCGAGAGTGTCTCAACTCTGGGATAGAGTGC 360
QY AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db GACTCTCAGGTACCTGCATCAACCCCTTAACTGGTGTGATGGCTGTCACTGCCCC 420
QY GlyGlyGluAspGluAsnArgCysValArgLeuTrpGlyProAsnPheIleLeuGlnVal 160
Db GCGGGGAGGAGCGAAGTGGTGTGCTGCTCTACGGATCAAACTTCATCCTTCAGGTG 480
QY TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db TACTCATCTCAGAGGAGTCTGGCACCTGTGTGCCAAGACGACTGGACAGGAACCTAC 540
QY GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db GCGCGGGCGGCTGTACAGGACATGGGCTAAGAATAATTTTACTCTAGCCCAAGGAATA 600
QY ValAspAspSerGlySerThrSerPheMetLysLeuLeuAsnThrSerAlaGlyAsnValAsp 220
Db GTGGATGACAGGGATCCACGAGCTTTATGAACCTGAACCAAGTGGCGGCATGTGAT 660
QY IleTyrLysLeuLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db ATCTATAAAAACTGTACACAGTATGCTCTTCTTCAAAGCAGTGGTTTCTTTACGC 720
QY CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
Db TGTATAGCTCGGGGGTCACTTGAACCTCAAGCCCGCAGACGAGTGTGGCGCGAG 780
QY SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db AGCGGCTCCCGGGGGCTGGCCCTGGCAGTGCAGCTGCAGCGGCCACTGCGTGGAAAA 900
QY CysGlyGlySerIleIleTrpProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db TCGGGAGGCTCCATCATCAACCCCGAGTGGATCGTGACAGCGGCCACTGCGTGGAAAA 960
QY ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db CCTCTTAAACAATCCATGGCATTGGACGGCATTTGGGGGATTTTGGACAAATCTTTTCATG 960
QY PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db TTCTATGGAGCGGATACCAAGTAGAATAAGTATTTCTCATCCAAATTTATGACTCCAG 1020
QY ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db ACCAAGAACAATGACATTTGGCTGATGAAGTGCAGAGCCCTCTGACTTTCAACGACCTA 1080
QY ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db GTGAACCCAGTGTGTCTGCCCAACCCAGGCATGATGTGTCAGCCAGACAGCTCTGCTGG 1140
QY IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db ATTTCCGGGTGGGGGGCCACCGAGGAGGAAGGAGACCTCAGAAAGTGTGAACGCTGCC 1200
QY LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db AAGGTGCTTCTCATTTGACACACAGATGCAACAGCAGATATGTCTATGACACCTGATC 1260
QY ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db ACACGAGCATGATCTGTGGGCTTCTGTGAGGGGAAACGTCGATTTCTTCCAGGGTGAC 1320

QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGGCTCTGCTGCTCACTTCGAAGAACAAATATCTGCTGGCTGATAGGGATACAAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGGTTCTGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGGTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTTATCGACAAATGAGGGGACAGCGC 1476

Search completed: June 1, 2004, 19:17:39
Job time : 749 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 1, 2004; 14:46:00 ; Search time 131 Seconds
(without alignments)
2084.243 Million cell updates/sec

Title: US-09-615-285B-2

Perfect score: 2717

Sequence: 1 MALNSGSPPAIGPYENHGY.....VYGNVWFTDIYRQMRADG 492

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US09615285/runat_01062004_132134_19976/app_query.fasta_1.647
-DB=issued Patents NA -QFMT=fastext -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09615285@cgn2_1 69 @runat_01062004_132134_19976 -NCPUS=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/ECTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	1479	US-09-342-749-1	Sequence 1, Appli
2	2717	100.0	1479	US-09-691-840-1	Sequence 1, Appli
3	2696	99.2	2479	US-09-342-749-29	Sequence 29, Appl
4	2696	99.2	2479	US-09-691-840-29	Sequence 29, Appl
5	2696	99.2	2479	US-09-685-166A-894	Sequence 894, App
6	1569	57.7	1077	US-08-807-151-2	Sequence 2, Appli
7	1569	57.7	1077	US-09-478-957-2	Sequence 2, Appli
8	1484	54.6	1001	US-09-016-434-963	Sequence 963, App
9	1165	42.9	693	US-09-685-166A-896	Sequence 896, App
10	888	32.7	2413	US-09-518-046-1	Sequence 1, Appli
11	863.5	31.8	2544	US-09-518-046-3	Sequence 3, Appli
12	858	31.6	2416	US-09-261-416-1	Sequence 1, Appli

13	687.5	25.3	2079	4	US-09-656-002-1	Sequence 1, Appli
14	687.5	25.3	2079	4	US-09-851-588-5	Sequence 5, Appli
15	2038	24.9	2038	3	US-09-008-271A-18	Sequence 18, Appli
16	676.5	24.9	2081	4	US-09-851-588-7	Sequence 7, Appli
17	660	24.3	2581	1	US-08-200-900A-1	Sequence 1, Appli
18	660	24.3	2581	5	PCT-US94-00616-1	Sequence 1, Appli
19	590.5	21.7	1739	2	US-08-681-151-2	Sequence 2, Appli
20	585.5	21.5	2363	4	US-09-742-703-3	Sequence 3, Appli
21	580	21.3	1783	3	US-09-510-738A-188	Sequence 188, App
22	580	21.3	1783	4	US-09-861-966-188	Sequence 188, App
23	558.5	20.6	3147	2	US-09-027-337-1	Sequence 1, Appli
24	558.5	20.6	3147	4	US-09-644-600-1	Sequence 1, Appli
25	558.5	20.6	3147	4	US-09-644-600-18	Sequence 18, Appli
26	558.5	20.6	3147	4	US-09-654-600A-1	Sequence 1, Appli
27	558.5	20.6	3147	4	US-09-654-600A-18	Sequence 18, Appli
28	557	20.5	1515	4	US-09-820-002-1	Sequence 1, Appli
29	554.5	20.4	2152	4	US-09-023-655-157	Sequence 157, App
30	540.5	19.9	1605	2	US-09-008-846-1	Sequence 1, Appli
31	527.5	19.4	1460	4	US-09-370-838-80	Sequence 80, Appli
32	527.5	19.4	1462	4	US-09-370-838-55	Sequence 55, Appli
33	527.5	19.4	1517	1	US-08-508-448C-15	Sequence 15, Appli
34	527.5	19.4	2790	4	US-09-370-838-79	Sequence 79, Appli
35	513.5	18.9	696	1	US-08-508-448C-24	Sequence 24, Appli
36	500.5	18.4	959	4	US-09-023-942A-25	Sequence 25, Appli
37	497.5	18.3	901	1	US-08-508-448C-9	Sequence 9, Appli
38	497.5	18.3	2900	2	US-09-027-337-9	Sequence 9, Appli
39	497.5	18.3	2900	4	US-09-644-600-9	Sequence 9, Appli
40	497.5	18.3	2900	4	US-09-654-600A-9	Sequence 9, Appli
41	490	18.0	1553	4	US-09-280-116-10	Sequence 10, Appli
42	485.5	17.9	933	4	US-09-023-942A-29	Sequence 29, Appli
43	475	17.5	1225	4	US-09-734-675-1	Sequence 1, Appli
44	471	17.3	1110	4	US-09-386-653A-1	Sequence 1, Appli
45	468.5	17.2	980	4	US-09-023-942A-30	Sequence 30, Appli

ALIGNMENTS

RESULT 1

US-09-342-749-1
; Sequence 1, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1476)
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (724)
; OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (985)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1347)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329

```

FEATURE:
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329.
FEATURE:
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
FEATURE:
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
FEATURE:
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to
OTHER INFORMATION: Ile
US-09-342-749-1

Alignment Scores:
Pred. No.:          9,21e-258          Length:          1479
Score:              2717.00             Matches:          492
Percent Similarity: 100.00%             Conservative:    0
Best Local Similarity: 100.00%           Mismatches:      0
Query Match:        100.00%             Indels:          0
DB:                  3                   Gaps:            0

US-09-615-285B-2 (1-492) x US-09-342-749-1 (1-1479)

Qy  1  MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr  20
Db  1  ATGCTTTTGAACTCAGGGTCACCCAGCAGCTATTGGACCTTACTATGAAAACCATGGATAC  60

Qy  21  GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis  40
Db  61  CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGGTCCCCACCTGTCTACAGGTGCAT  120

Qy  41  ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla  60
Db  121  CCGGCTCAGTACTACCCGTCCTCCCGTGGCCAGTACGCCCGAGGGTCTCTGACGCGAGCT  180

Qy  61  SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys  80
Db  181  TCCAAACCCGCTGCTCAGCGAGCCCAATCCCAATCCCGACAGTGTGACCTTCAAG  240

Qy  81  ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu  100
Db  241  ACTAAGAAGACCTGTGCATCAGCTTGAACCTTGGAGCCCTTCTCTGTGGAGCTGCGCTG  300

Qy  101  AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys  120
Db  301  GCCGCTGGCTACTCTGGAAGTTCATGGCAGCAGCAAGTGTCCAACTCTGGGATAGATGC  360

Qy  121  AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro  140

```

```

Db  361  GACTCCTCAGTACTCTGCATCAACCCCTCTAACTGGTGTGATGGCTGCACACTCCCC  420
Qy  141  GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal  160
Db  421  GCGCGGAGGAGCAGAAATCGGTGTGTTCGGCTCTACGACCAAACTTCATCTCTCAGGTG  480
Qy  161  TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr  180
Db  481  TACTCATCTCAGAGGAGTCTTGGACCCCTGTGTCCAGACGACCTGGACGAGAACTAC  540
Qy  181  GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle  200
Db  541  GCGCGGCGGCTCTGACGGACATGGGCTATAAAGATAATTTTACTCTAGCCCAAGAAATA  600
Qy  201  ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp  220
Db  601  GTGGATGACACGGATCCACCGCTTTATGAACCTGAACACAGTGGCGGCAATGTTCGAT  660
Qy  221  IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg  240
Db  661  ATCTATAAAAACTGTACCAAGTGTCTCTTCAAAAGCAGTGGTCTTTTTCACGC  720
Qy  241  CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu  260
Db  721  TGTATAGCTCGGGGTCACTTGAACCTCAAGCCGCCAGACAGGATCGTGGCGCGAG  780
Qy  261  SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal  280
Db  781  AGCGCGCTCCCGGGGCTGGCCCTGGCAGGTGAGCTGCACGTGACAGACGTCCACGTG  840
Qy  281  CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys  300
Db  841  TCGGAGGCTCCATCATCATCCCCCGAGTGGATCGTGACAGCCGCCACTGCGTGGAAAA  900
Qy  301  ProLeuAsnAsnProTyrHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet  320
Db  901  CCTCTTAAACATCCATGGCATTTGGACGGCAATTTGCGGGATTTTGAGACAACTTTTC  360
Qy  321  PheTyrGlyAlaGlyTyrGlnValGlyValIleSerHisProAsnTyrAspSerLys  340
Db  961  TTCTATGAGCGCGATCCAAAGTAGAAAAAGTATTCTCATCCAAATATGACTCCAAG  1020
Qy  341  ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu  360
Db  1021  ACCAAGAAATGACATTTGCTGATGAGCTGAGAGCTCTGACTTTCACAGACCTA  1080
Qy  361  ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp  380
Db  1081  GTGAAACCAAGTGTCTGCTGCCCAACCCAGGATGATGTGCAGCCAGACAGCTCTGCTG  1140
Qy  381  IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla  400
Db  1141  ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGACGCTGCC  1200
Qy  401  LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle  420
Db  1201  AAGGTGCTTCTCATTTGACACACAGATGCAACAGCAGATATGTCTATGACAACTGATC  1260
Qy  421  ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp  440
Db  1261  ACACCCAGCATGATCTGTGCGGCTTCTTGCAGGGGACCGTGCATCTTGTGCCAGGCTAC  1320
Qy  441  SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer  460
Db  1321  AGTGAGGGCTCTGGTCACTTCGAAGAACAAATATCTGGTGGCTGATAGGGATACAGC  1380
Qy  461  TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe  480
Db  1381  TGGGGTCTGGCTGTGCCAAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTATTTC  1440
Qy  481  ThrAspTrpIleTyrArgGlnMetArgAlaAspGly  492

```

Db 1441 ACAGCTGGATTATCGACAAATGAGGGCAGACGGC 1476

RESULT 2

US-09-691-840-1
 ; Sequence 1, Application US/09691840
 ; Patent No. 6444419
 ; GENERAL INFORMATION:
 ; APPLICANT: Wong, Alexander K.C.
 ; APPLICANT: Tavtigian, Sean V.
 ; APPLICANT: Teng, David H.-F.
 ; APPLICANT: Myriad Genetics, Inc.
 ; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
 ; FILE OF INVENTION: 2318-202
 ; CURRENT APPLICATION NUMBER: US/09/691,840
 ; CURRENT FILING DATE: 2000-10-18
 ; PRIOR APPLICATION NUMBER: US/09/342,749
 ; PRIOR FILING DATE: 1999-06-29
 ; PRIOR APPLICATION NUMBER: US 60/091,044
 ; PRIOR FILING DATE: 1998-06-29
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1479
 ; TYPE: DNA

ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1476)
 ; NAME/KEY: conflict
 ; LOCATION: (724)
 ; OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
 ; NAME/KEY: conflict
 ; LOCATION: (985)
 ; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: conflict
 ; LOCATION: (1347)
 ; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: conflict
 ; LOCATION: (1466)
 ; OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: conflict
 ; LOCATION: (1471)
 ; OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: allele
 ; LOCATION: (478)
 ; OTHER INFORMATION: This base can be G or A with G being the more common allele. The codon will change from Val to Met.
 ; NAME/KEY: allele
 ; LOCATION: (777)
 ; OTHER INFORMATION: This base can be C or T with C being the more common allele. The codon is unaffected with both alleles encoding Gly.
 ; NAME/KEY: allele
 ; LOCATION: (768)
 ; OTHER INFORMATION: This base can be C or T with C being the more common allele. This is a silent polymorphism.
 ; NAME/KEY: allele
 ; LOCATION: (834)
 ; OTHER INFORMATION: This base can be C or T with C being the more common allele. This is a silent polymorphism.
 ; NAME/KEY: allele
 ; LOCATION: (625)
 ; OTHER INFORMATION: This base can be T or A with T being the more common allele. The codon will change from Phe to Ile
 ; OTHER INFORMATION: common allele. The codon will change from Phe to Ile

US-09-691-840-1
 Alignment Scores:
 Pred. No.: 9 21e-258 Length: 1479
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0

DB: 4 Gaps: 0

US-09-615-285b-2 (1-492) x US-09-691-840-1 (1-1479)

QY	1	MetAlaLeuAenSerGlySerProAlaIleGlyProTyrTyrGluAenHisGlyTyr	20
DB	1	ATGGCTTTGAACCTCAGGCTCAGGCTCAGGCTTACTATGAAACCATGGATAC	60
QY	21	GlnProGluAenProTyrProAlaGlnProThrValProThrValTyrGluValHis	40
DB	61	CAACCGGAAACCCCTATCCCGCAGCAGCCACCTGTGTGCTCCCTCTCTACGAGGTGCAT	120
QY	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
DB	121	CGGCTCAGTACTACCGCTCCCGTCCCGCAGTACGCCCCCGAGGCTCTGACGAGGCT	180
QY	61	SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
DB	181	TCCAAACCCCGCTGCTGCGACGACCCAAATCCCATCCCGGACAGTGTGCACCTCAAG	240
QY	81	ThrLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu	100
DB	241	ACTAAGAAAGCACCTGTGCATCACCTTGACCTGGGACCTTCCTCGTGGAGCTGGCTG	300
QY	101	AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAenSerGlyIleGluCys	120
DB	301	GGCGTGGCTTACTCTGAAAGTTCATGGCAGCAGAGTCTCCAACTCTGGATAGAGTGC	360
QY	121	RepSerSerGlyThrCysIleAenProSerAenTyrCysAenGlyValSerHisCysPro	140
DB	361	GACTCCTCAGGTACTCTGCATCAACCTCTAACTGTGTGTGATGGCTGTGCACACTGCCCT	420
QY	141	GlyGlyGluAenGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal	160
DB	421	GGCGGGGAGGAGGAGATCGGTGTGCTCTACGGACCAACTTCATCTTCAGGTG	480
QY	161	TyrSerSerGlnArgLysSerTyrHisProValCysGlnAenSerTyrPheAenGluSer	180
DB	481	TACTCATCTCAGAGGAGTCTCGGACCTCTGTGCAAGACGACTCGGAGACGAGACTAC	540
QY	181	GlyArgAlaAlaCysArgAenMetGlyTyrLysAenAenPheTyrSerSerGlnGlyIle	200
DB	541	GGCGGGGCGGCTCAGGAGCATGGGCTATAAGATTAATTTTACTCTAGCCAGGAGATA	600
QY	201	ValAenSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAen	220
DB	601	GTGGATCAGACGGGATCCACGAGCTTTATGAAACTGAACACAGTCCCGGCAATGTTCGAT	660
QY	221	IleTyrLysLysLeuTyrHisSerAenAlaCysSerSerLysAlaValValSerLeuArg	240
DB	661	ATCTATAAAGAACTGTACACAGTGTGTCTTCTTCAAGGACAGTGTCTTTTACGC	720
QY	241	CysIleAlaCysGlyValAenLeuAenSerArgGlnSerArgIleValGlyGlyGlu	260
DB	721	TGTATAGCTTGGGGGTCAACTTGAACCTCAAGCCCGCAGCAGGATCTGGGGCGGAG	780
QY	261	SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAenValHisVal	280
DB	781	AGCGGCTCCCGGGGCTCGCTCGCTGGAGTCTGAGCTCAGCTCCAGAACCTCCACGCTG	840
QY	281	CysGlyGlySerIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys	300
DB	841	TGGGAGGCTCATCATCACCCCGAGTGTGATCGTACAGCCCGCCCTCTGCTGGAGAAA	900
QY	301	ProLeuAenAenProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
DB	901	CCTCTTAACATCCATGGCATTTGACGGCATTTTGGGGGATTTTGAGACAACTTCTCATG	960
QY	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAenTyrAenSerLys	340
DB	961	TTCTATGAGCCGGATACCAAGTAGAAGATGATTTCTATCCCAATTTATGACTCCAG	1020
QY	341	ThrLysAenAenAenPheIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu	360

Db	1021	ACCAAGAACAAATGACATTGGCGCTGATGAAGCTCCAGAGCCTCTGACTTTCACGACCTA	1080
Qy	361	VallYsProValCYsLeuProAsnProGlyWe-MetLeuGlnProGluGlnLeuCYsTrp	380
Db	1081	GTGAACACAGTGTGTCTCCCAACCCAGGCATGATCTGCAGCCAGACAGCTCTGTCTGG	1140
Qy	381	IleSerGlyTTPGlyAlaThrGluGluIuYsGlyThrSerGluValLeuAsnAlaAla	400
Db	1141	ATTTCCGGTGGGGGGCCACCAGGAGGAAGAGGAGACCTTCAGAGTGTCTGAAGCGTGCC	1200
Qy	401	LysValLeuLeuIleGluThrGlnArgCYeAsnSerArgTyrValTyrAspAsnLeuIle	420
Db	1201	AAGGTGCTCTCATTTGAGACACAGAGATGCAACAGCAGACATATGCTATGACAACTGATC	1260
Qy	421	ThrProAlaMetIleCYeAlaGlyPheLeuGlnGlyAsnValAsnSerCYeGlnGlyAsp	440
Db	1261	ACACAGCGCCATGATCTGTGCCGGCTTCCTCAGGGGAAGCTGATTCATGCCAGGGTGAC	1320
Qy	441	SerGlyGlyProLeuValThrSerLYeAsnAsnIleTTPLeuIleGlyAspThrSer	460
Db	1321	AGTGAGGGCCCTCTGGTCACCTTCGAAGAACAAATATCTGGTGGCTGATAGGGGATACAAGC	1380
Qy	461	TPGlySerGlyCYeAlaLYeAlaTyrArgProGlyValTyrGlyAsnValMetValPhe	480
Db	1381	TGGGGTCTTGACTGTGCCAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGATTC	1440
Qy	481	ThrAspTTPileTyrArgGlnMetArgAlaAspGly	492
Db	1441	ACGGACTCGATTATTCACAAATCAGGGCAGACGGC	1476

RESULTS

US-09-342-749-29
; Sequence 29, Application US/09342749
; Patent No. 6166194

```

/ GENERAL INFORMATION:
/ APPLICANT: Wong, Alexander K.C.
/ APPLICANT: Tavtigian, Sean V.
/ APPLICANT: Feng, David H.-F.
/ APPLICANT: Myriad Genetics, Inc.
/ TITLE OF INVENTION: TWPGRS2 is a Tumor Suppressor
/ FILE REFERENCE: 2318-202
/ CURRENT APPLICATION NUMBER: US/09/342,749
/ CURRENT FILING DATE: 1999-06-29
/ EARLIER APPLICATION NUMBER: US 60/091,044
/ EARLIER FILING DATE: 1998-06-29
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: Patent Ver. 2.0
/ SEQ ID NO 29
/ LENGTH: 2479
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-342-749-29

```

Alignment Scores:		
Pred. No.:	2.42e-355	Length: 2479
Score:	2696.00	Matches: 486
Percent Similarity:	99.60%	Conservative: 5
Best Local Similarity:	98.78%	Mismatches: 1
Query Match:	99.23%	Indels: 0
DB:	3	Gaps: 0

US-09-615-285B-2 (1-492) X US-09-342-749-29 (1-2479)

1	MetAlaIeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
57	ATGGCTTTGAACCTACAGGCTCACCCACAGCTATTGGACCTTACTATGAACCACTGGATAC	116
21	GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis	40
117	CAACCCGAAAAACCCCTATTCCCGCACACCCACCTGTGTGCCCATCTGTCTACGAGGTGCAT	176
41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValIleThrGlnAla	60

Db 1257 AAGGTGCTTCTCATTGACACACAGAGATGCAACACAGATATGTCTATGACCAACCTGATC 1316
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyVasp 440
Db 1317 ACACAGCCATGATGCTGCGGGTCTCTCGAGGGGACGTCGATTTCTTGCCAGGGTGAC 1376
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1377 AGTGAGGGCCTCTGGTCACATCGAACCAACAATATCTGGTGGCTGATAGGGGATACAAGC 1436
Qy 461 TrpGlySerGlyCysAlaGlyAlaValArgProGlyValTrpGlyAsnValMetValPhe 480
Db 1437 TGGGGTCTGGCTGTGGCCAAAGCTTACAGACAGAGGTACGGGAATGTGATGTTTC 1496
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1497 ACGGACTGATTTATCGACAAATGAAGCAACAGGC 1532

RESULT 4
US-09-691-840-29
; Sequence 29, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TSPRS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-691-840-29

Alignment Scores:
Pred. No.: 2,428-255 Length: 2479
Score: 2696.00 Matches: 486
Percent Similarity: 99.80% Conservative: 5
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 99.23% Indels: 0
DB: Gaps: 0

US-09-615-285B-2 (1-492) x US-09-691-840-29 (1-2479)

Qy 1 MetaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 57 ATGGCTTTGAATCACTGAGGTCAACACAGCTATTGGACCTTACTATGAAACCATGGATAC 116
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 117 CAACCGGAACAAACCTTATCCCGCAGACGCCACTGTGGTCCCACTGTCTACAGGTGCA 176
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 177 CCGGCTCAGTACTACCGCTCCCGCTGGCCAGTAGTACGCCCGCGAGGCTCTGACGAGGCT 236
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 237 TCCAAACCCGCTGCTGACGAGCCCAATCCCATCCGAGACAGTGTGACCTCAAG 296
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 297 ACTAGAAAGCACTGTGCACTACCTTGACCCCTGGGACCTTCTCTGTGGAGCTGGCTG 356

Qy 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 357 GCGCTGGCCTACTCTGGAAAGTTCTGGGCAGCAAGTGTCTCCAACTCTCGGATAGAGTGC 416
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 417 GACTCTCTAGGTACCTGTCATCAACCCCTCTAATCGTGTGATGGCGTGTACACTGCCCC 476
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 477 GCGGGGAGGAGAGAAATCGGTGTGTTCGCTCTACGGACCAAACTTCTATCTCTCAGATG 536
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 537 TACTCATCTCAGAGGAAGTCTGGCACCTCTGTGCGCAAGACACTGGAAACGAACTAC 596
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 597 GCGCGGGCGGCTGCAGGACATGGCTATAGAAATAATTTTACTCTAGCCAAAGAAATA 656
Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 657 GTGGATGACAGCGGATCCACGACTTATGAAACTGACACAAAGTCCCGCAATGTCGAT 716
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 717 ATCTATAAATACTGTACACAGTGTGCTGTCTTCAAGACAGTGGTCTTTTACGC 776
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 777 TGTTTAGCCTGCGGGTCAACTTGAACCTCAAGCGCGCAGCAGGAGTCTGGGCGGTGAG 836
Qy 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 837 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGTCAAGCTGCACGCTCAGAACGTCACGTC 896
Qy 281 CysGlyCysIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 897 TGGGAGGCTCCATCATCAACCCGAGTGGATCGTGACAGCCGCCCACTGCGTGGAAAA 956
Qy 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 957 CCTCTTAACAATCCATGGCATTTGGCGGCATTTGCGGGGATTTTGAGACATCTTTTCATG 1016
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1017 TTCTATGGAGCCGATACCAAGTACAAAAAGTATTTCTCATCCAAATATGACTCCAAAG 1076
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1077 ACCAAGAACAAATGACATTTGGCTGATGAAGCTGCAGAAAGCTCTGACTTTCACAGCCTA 1136
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1137 GTGAACACAGTGTCTGCCAACCCAGCGATGATGCTGCAGCCAGACAGACTCTGCTGG 1196
Qy 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1197 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAGAACCTCAGAGTGTGGAACGCTGCC 1256
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAsnLeuIle 420
Db 1257 AAGGTGCTCTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGCAACCTGATC 1316
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1317 ACACAGCCATGATCTGTGCGGCTTCTCTGACGGGAACGTCGATTTCTTGCCAGGGTGAC 1376
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1377 AGTGAGGGCCTCTGGTCACTTCGAAACAAATATCTGGTGGCTGTAGGGGATACAAGC 1436
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480

203	TGCGGGTCAACTTGA	CTCAAGCCCGCAGAGCAGGATCGTGGCGCGCAGAGCGCGCT	266
264	ProGlyAlaTrpProIn	ValSerLeuHisValGlnAsnValHisValCysGlyGly	283
263	CGGGGGCTGGCCCT	GGCAGGTGACGCTCCAGAACGTCACGCTGTGGCGAGCG	322
284	SerIleIleThrProGlu	TrpIleValThrAlaAlaHisCysValGluIysProLeuAsn	303
323	TCCATCATCACCCCG	AGTGGATCGTGACAGCCGCCCACTGGGTGGAAAAACCTCTTAAC	382
304	AsnProTrpHisTrpThr	AlaPheAlaGlyIleLeuArgGlnSerPheMetPheTrpGly	323
383	AATCATGGCATTGGAC	GGCATTTGGGGGATTTTCAGACAATCTCTTCATGTTCTATGGA	442
324	AlaGlyTyzGlnValGlu	IysValIleSerHisProAsnTyrAspSerIysThrIysAsn	343
443	CGCGGATACCAAGTA	GAAGATGATTCATCCAAATATGACTCCAAAGACCCAGAAC	502
344	AsnAspIleAlaLeuMet	IysLeuGlnIysProLeuThrPheAsnAspLeuValIysPro	363
503	AATGACATTGGCTGAT	GAAGCTGCAGAGCCCTCTGACTTTCAACGACCTAGTGTGAACCA	562
364	ValCysLeuProAnProGly	MetMetLeuGlnProGluGlnLeuCysTrpIleSerGly	383
563	GTGTGTCTGGCCCAAC	CCAGCGCATGCTGCACCCAGACAGCTCTGCTGGGATTTCCGGG	622
384	TrpGlyAlaThrGluGlu	IysGlyIysThrSerGluValLeuAsnAlaAlaIysValLeu	403
623	TGGGGGCCCCAGGAG	GAAGGAGACCTCAGAAAGTCTGAAAGCTGCCAAGGTGCTT	682
404	LeuIleGluThrGlnArg	CysAsnSerArgTyrValTyrAspAsnLeuIleThrProAla	423
683	CTCATTGAGACACAGA	TATCAACAGCAGAGATGTCTATGACAACTGATCACACGACC	742
424	MetIleCysAlaGlyPhe	LeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGly	443
743	ATGATCTGTGCGCGCT	TTCCTGACGGGACCTCGATCTTTGCCAGGGTGACAGTGGAGGG	802
444	ProLeuValThrSerIys	AsnAsnIleTrpTrpLeuIleGlyAspThrSerTrpGlySer	463
803	CNTCTGGTCACTTCGA	AGAACCAATATCTGTGTGCTGATAGGGGATACAAGCTGGGGTCTCT	862
464	GlyCysAlaIysAlaTyr	ArgProGlyValTyrGlyAsnValMetValPheThrAspTrp	483
863	GGCTGTGCCAAGCTTC	ACAGCATACAGACAGGNGTGTACGGGAATGTGATGTATTACAGGACTGG	922
484	IleTyrArgGlnMetArg	AlaAspGly	492
923	ATTTATGACAAATGAG	CGCAGACGGC	949

```

US-09-478-957-2
; Sequence 2, Application US/09478957
; Patent No. 6350448
;
; GENERAL INFORMATION:
;
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
;
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
;
; NUMBER OF SEQUENCES: 5
;
; CORRESPONDENCE ADDRESS:
; ADDRESSSE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
;
; COUNTRY: US
; ZIP: 94304
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
;
; CURRENT APPLICATION DATA:

```

RESULT 6

323 TCCATCATCACCCCCGAGTGGATCGTGACAGCGCCCACTGCGTGGAAAAACCTCTTAC 382

Qy 304 AsnProT¹HisT¹ThrAlaPheAlaGlyIleIeuArgGlnSerPheMet¹heTyrGly 323

D_b 383 AATTCATGGCATTTGGACGGCATTTGCGGGGATTTTGGACAAATCTTTTCATGTTCTATGGA 442

Qy 324 AAGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAsn 343

Db 443 GCCGGATACCAAGTAGAAAAAGTGATTCTCATCCAAATTATGACTCCAGACCAAGAAC 502

Qy 344 AsnAspIleAlaLeuMetIysLeuGlnLysProLeuThrPheAsnAspLeuValLysPro 363

Db 503 AATGACATTGGCGCTGATGAAGCTGCAGAAGCCTCTGACTTTCACAGCACTAGTGAAACCA 567

364 ValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGly 38

Db 563 GTGTGCTGCCCAACCCAGGATGATGTGTGCAGCAGACAGCTCTGTCTGGAATTTCCGGG 622

384 TrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeu 400

Db 623 TGGGGGGCCACCGAGGAGAAGGGGAAGACCTCAGAAGTGTCTGAACGCTGCCAAGGTGCTT 688

Qy	404	Leu	Ile	Leu	Thr	Gln	Arg	Cys	Asn	Ser	Arg	Val	Val	Asp	Asn	Leu	Leu	Pro	Ala	422
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

683 CTGATTGAGACACAGAGATGCAACAGCAGAGATATGTCTATGACAAACCTTGATCACACCCAGCC / 443

[illegible]

DB	743	ATGATCTGTGCCCGCTTCCTCGAGGGGAACGTCGATTCTTCCTCAGGGTGATAGATGGAGGG
Qu	444	ProfenValThrSerIleAsnIleProTrpLeuIleAlaValThrSerTrpGlyValSer 463

444	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
445	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
446	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
447	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
448	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
449	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
450	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
451	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
452	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
453	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
454	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
455	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
456	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
457	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
458	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
459	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
460	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
461	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e	y	a	s	p	i	m	u	c	i	n	p	o	r	
462	Protein	val	ser	ly	sas	as	u	u	l	e	r	p	u	e	n	e	r	e														

[illegible]

863 GGCTGTGCCAAGACCTTACAGACCAGGAGTGTA CCGGAATGTGATGTGTATTTCACGGACTGG 922

Qv 484 IleTyrArgGlnMetArgAlaaspGly 492

923 ATTATATCGACAAATGAGGGCAGACGGC 949

RESULT 7

US-09-478-957-2
; Sequence 2, Application US/09478957

Patent No. 6350448 ; GENERAL INFORMATION:

APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti

1	TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
2	
3	TITLE OF INVENTION: PROTEASE
4	

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

;; ADDRESS: Incyte Pharmaceuticals, Inc.
;; STREET: 3174 Porter Drive

CITY: Palo Alto
STATE: CA

COUNTRY: US
ZIP: 94304

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; RECORDS PER TRACK: 80
; TRACKS PER SIDE: 6
; SIDES: 2
; COPIES OF EACH SIDE: 1
```

```

;
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
COMMUNIT
E-Link Version: 0.0
;

```

```

; SOFTWARE: Fastseq version 2.0
;
; CURRENT APPLICATION DATA:
;

```

```

; APPLICATION NUMBER: US/09/478,957
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,151
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,748
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORN0T01
; CLONE: 556016
; US-09-478-957-2

Alignment Scores:
Pred. No.: 5,78e-145 Length: 1077
Score: 1569.00 Matches: 288
Percent Similarity: 99.65% Conservativeness: 0
Best Local Similarity: 99.65% Mismatches: 1
Query Match: 57.75% Indels: 0
DB: 4 Gaps: 0

US-09-615-285b-2 (1-492) x US-09-478-957-2 (1-1077)
QY 204 SerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAspIleTyrLys 223
DB 83 AGCGATCCACAGCTTTATGAACCTGACACAGAGCGCGGCAATGCTATCTATATAA 142
QY 224 LysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArgCysIleAla 243
DB 143 AAACGTGACACAGTGTGCTCTCTTCAAAAGCAGTGTGTTCTTACGCTGTATAGC 202
QY 244 CysGlyValAsnLeuAsnSerArgGlnSerArgIleValGlyGlyGluSerAlaLeu 263
DB 203 TCGGGGTCAACTTGAACTCAACCCGACAGCAGGATGTGGGGCGGAGAGCGGCTC 262
QY 264 ProGlyValTrpProTrpGlnValSerLeuHisValGlnAsnValHisValCysGlyGly 283
DB 263 CCGGGGCGCTGGCCCTGGCAGGTGAGCTGCAGTCCAGAACGTCACCGTGTGGGAGGC 322
QY 284 SerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLysProLeuAsn 303
DB 323 TCCATCATCACCCTCCAGTGGATCGTGACAGCCGCCCACTGCGTGAAGAAACCTCTTAA 382
QY 304 AsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGly 323
DB 383 AATCCATGGCATTTGGAGCGGATTTGGGGGATTTTGAGACAATCTTTCATGTTCTATGA 442
QY 324 AlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAsn 343
DB 443 GCGGGATACCAAGTAGAAAAAGTATTCTCATCCAAATTTATGACTCCAAGACCAAGAAC 502
QY 344 AsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysPro 363
DB 503 AATGACATTCGGCTGATGAAGCTGCAGAGCCCTCTGACTTTCACACCTAGTGAACCA 562
QY 364 ValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpLysSerGly 383
DB 563 GTGTGTCTGCCCAACCCAGGCGATGATGCTGCAGCCAGAACAGCTCTGCTGGATTTCCGGG 622
QY 384 TrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaLysValLeu 403
DB 623 TGGGGGGCCCAACCGAGGAGAAAGGAGACCTCGAAGTGTCTGAACGCTGCCAAGGTGCTT 682
```

```

QY 404 LeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAla 423
DB 683 CTCATTGAGACACAGATGCAACAGCAGATATGCTATGACACACCTGATCACCAGCC 742
QY 424 MetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGly 443
DB 743 ATGATCTGTGCGGCTTCTTCCAGGGGAAAGCTCGATTTCTGCCAGGCTGACAGTGGAGG 802
QY 444 ProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSerTrpGlySer 463
DB 803 CNTCTGTGCTACTCGAAGAACAAATATCTGTGGTGGCTGATAGGGGATACAAAGTGGGGTCT 862
QY 464 GlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPheThrAspTrp 483
DB 863 GGCTGTGCCAAAGCTTACAGACACAGGAGTGTACGGGAATGTGATGTATTACCGACTGG 922
QY 484 IleTyrArgGlnMetArgAlaAspGly 492
DB 923 ATTTATCGACAAATGAGGCGACGCGC 949

RESULT 8
US-09-016-434-963
; Sequence 963, Application US/09016434
; Patent No. 650938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 963:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT05
; CLONE: 842889
; US-09-016-434-963

Alignment Scores:
Pred. No.: 1.2e-136 Length: 1001
Score: 1484.00 Matches: 279
Percent Similarity: 94.61% Conservativeness: 2
```

Best Local Similarity: 93.94%		Mismatches: 12	
Query Match: 54.62%		Indels: 4	
DB: 4		Gaps: 0	
US-09-615-285B-2 (1-492) x US-09-016-434-963 (1-1001)			
QY	1	MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
DB	109	ATGGCTTTGAACCTCAGGCTCACCACCGCTATTGGACCTTACTATGAAACCACTGGATAC	168
QY	21	GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis	40
DB	169	CAACCGGAAACCCCTATCCCGCAGCCCACTGTGTCCTTACGAGGTGCAT	228
QY	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
DB	229	CCGCTCAGTACTACCGTCCCGCTGCCAGTACGCCCGAGGCTCTTACGCGAGCT	288
QY	61	SerAsnProValValCysThrGlnProLysSerPro--SerGlyThrValCysThrSerL	80
DB	289	TCCAACCCGCTGCTGCAGCAGCCCAATCCCATCTCKGACNTGTGTGCACCTCAA	348
QY	80	ysThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyValAla	100
DB	349	AGACTAAGAAAGCACTGTGCATCACCTTGACCTTGGGAGCCTTCTCTGGAGCTGCC	408
QY	100	euAlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluC	120
DB	409	TGGCGCTGGCTACTCTGGAAGTTCATGGCAGCAAGTGTCTCAACTCTGGGATAGT	468
QY	120	ysAspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysP	140
DB	469	CGCACTCTCAGGTACCTGCATCAACCCCTCTAAGTGTGTGGTGTGCACATGCC	528
QY	140	roGlyGlyGlu-AspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGln	159
DB	529	CCGCGGGGAGGACAARAATCGGTGTTTTCCTCTACGGACCAAACTTCATCTTCAG	588
QY	160	ValTyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAsnGluAsn	179
DB	589	GTGTACTCATCTCAGAGAACTCTCGCACCCCTGTGTGCCAAGACNACTGGAACGAA	648
QY	180	TyrGlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGly	199
DB	649	TACGGCGGGCGGCTCGCAGGACATGGCTATAAGATTAATTTTACTCTAGCCAAAG	708
QY	200	IleValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnVal	219
DB	709	ATAGTGGATGACAGCGGATCNCAGCTTTATGAACTGAACACACAGTGCAGCAATGTC	768
QY	220	AspIleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeu	239
DB	769	GATATCTATAAAACCTGTACACAGTATGCCCTGTCTTCAAAAGCAGTGGTTCTTTA	828
QY	240	ArgCysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGly	259
DB	829	CGCTGTATACCTTGGGGGTCACTTGAACCTAAGCCGCGCAGACAGATCGTGGGGGC	888
QY	260	Glu-SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHi	279
DB	889	GAANAGCGGCTCCCGGGGCTCGCCCTGGCAGGTGAGCCTGCAGCTCCAGAACTCCAC	948
QY	279	sValCysGlyGlySerIleIleThrProGluTyrIleValThr	293
DB	949	NTTTTGCNAGAGCTCCATCATCAACCCCGAATTGATCGTGACA	991

RESULT 9

US-09-685-166A-896

: Sequence 896, Application US/09685166A

: Patent No. 6630305

: GENERAL INFORMATION:

: APPLICANT: Xu, Jiangchun

: APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Harlocker, Susan L.

APPLICANT: Jiang, Yuqi

APPLICANT: Henderson, Robert A.

APPLICANT: Kalos, Michael D.

APPLICANT: Fanger, Gary R.

APPLICANT: Retter, Marc W.

APPLICANT: Stolk, John A.

APPLICANT: Day, Craig H.

APPLICANT: Vedvick, Thomas S.

APPLICANT: Carter, Darrick

APPLICANT: Li, Samuel

APPLICANT: Wang, Aijun

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Hepler, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.427C21

CURRENT APPLICATION NUMBER: US/09/685,166A

CURRENT FILING DATE: 2000-10-10

NUMBER OF SEQ ID NOS: 898

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 896

LENGTH: 683

TYPE: DNA

ORGANISM: Homo sapiens

US-09-685-166A-896

Alignment Scores:

Pred. No.: 1,67e-105

Length: 683

Matches: 208

Conservative: 1

Percent Similarity: 100.00%

Mismatches: 0

Best Local Similarity: 99.52%

Indels: 0

Query Match: 42.88%

Gaps: 4

DB: 4

US-09-615-285B-2 (1-492) x US-09-685-166A-896 (1-683)

QY 1

MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr

20

DB 57

ATGGCTTTGAACCTCAGGCTCACCACCGCTTACTATGAAACCACTGGATAC

116

QY 21

GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis

40

DB 117

CAACCGGAAACCCCTATCCCGCAGCCCACTGTGTCCTCCACTCTACGAGGTGCAT

176

QY 41

ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla

60

DB 177

CCGGCTCAGTACTACCGTGTCCCGCTGCCAGTACGCCCGCGAGGCTCTTGACGCA

236

QY 61

SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys

80

DB 237

TCCAACCCCGTGTCTGCAGCAGCCCAATCCCATCCCGGACAGTGTGCACCTCAAAG

296

QY 81

ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu

100

DB 297

ACTAAGAAAGCACTGTGCATCACCTTGACCTGCGGACCTTCTCTGTTGGAGCTGCGCTG

356

QY 101

AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys

120

DB 357

CCCGCTGCGCTACTCTGGAAGTTTCACTGGGAGCAAGTGTCCCACTCTGGGATAGAGTGC

416

QY 121

AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro

140

DB 417

GACTCTCAGGTACTGCATCAACCCCTCTAACTGGTGTGATGTCGGGTGTCCACTGCCCC

476

QY 141

GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal

160

DB 477

GGCGGGAGGACGAGAAATCGGTGTGTTTCCGCTCTACGGAACCAACTTTCATCTTCAGATG

536

QY 161

TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAsnGluAsnTyr

180

DB 537

TACTCATCTCAGAGAGTCTCTGCAACCCCTGTGTGCCAAGACGACTGGAACGAGAACTAC

596

RESULT 9
US-09-685-166A-896
; Sequence 896, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C21
CURRENT APPLICATION NUMBER: US/09/685,166A
NUMBER OF SEQ ID NOS: 898
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 896
LENGTH: 683
TYPE: DNA
ORGANISM: Homo sapiens
US-09-685-166A-896

Alignment Scores:
Pred. No.: 1,67e-105 Length: 683
Score: 1165.00 Matches: 208
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.52% Mismatches: 0
Query Match: 42.88% Indels: 0
DB: 4 Gaps: 0

US-09-615-285B-2 (1-492) x US-09-685-166A-896 (1-683)

Qy	1	MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
Db	57	ATGGCTTTGAACCTCAGGCTCACCACCGCTATTGGACCTTACTATGAAACCACTGGATAC	116
Qy	21	GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis	40
Db	117	CAACCGGAAACCCCTATCCCGCAGCCCACTGTGTCCTTACGAGGTGCAT	176
Qy	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
Db	177	CCGCTCAGTACTACCGTCCCGCTGCCAGTACGCCCGAGGCTCTTACGCGAGCT	236
Qy	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
Db	237	TCCAACCCCGTGTCTGTCAGCAGCCCAATCCCATCCGCGACAGTGTGCACCTCAAAG	296
Qy	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu	100
Db	297	ACTAAGAAAGCACTGTGCATCACCTTGACCTTGGGAGCCTTCTCTGGAGCTGCGTG	356
Qy	101	AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
Db	357	GCGCTGCGCTACTCTGGAAGTTCATGGCAGCAAGTGTCTCCAACTCTGGATAGATGC	416
Qy	121	AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro	140
Db	417	GACTCTCAGGTACCTGCATCAACCCCTTAACCTGTGTGATGGGTGTGCACACTGCCCC	476
Qy	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal	160
Db	477	GGCGGGGAGACGAGAAATCCGTGTGTTGCCCTCTACGGACCAAACTTCATCTTCAGATG	536
Qy	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAsnGluAsnTyr	180
Db	537	TACTCATCTCAGAGGAAGTCTTGGCAGCCCTGTGTGCCAAGACGACTGGAAACGAGACTAC	596

QY 181 GlyArgAlaAlaCysArgSerMetGlyTyrLysAsnAsnPhetYrSerSerGlnGlyLe 200
Db 597 GGG 656
QY 201 ValAspAspSerGlySerThrSerPhe 209
Db 657 GTGGATGACAGCGGATCCACCGAGCTTT 683

RESULT 10

US-09-518-046-1
; Sequence 1, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CJP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 2413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 gene
US-09-518-046-1

Alignment Scores:

Pred. No.: 2,248-77 Length: 2413
Score: 888.00 Matches: 205
Percent Similarity: 53.06% Conservative: 64
Best Local Similarity: 40.43% Mismatches: 186
Query Match: 32.68% Indels: 52
DB: 3 Gaps: 15

US-09-615-285B-2 (1-492) x US-09-518-046-1 (1-2413)

QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrVal----- 36
Db 24 AAGCCAGTAACACTGTGGCTACTATCTCTCCGTGGTGCATCATATTTTGGGACTC 83
QY 37 -----TyrGluValHisProAlaGlnTyrTyrProSerProValProGlnTyrAlaPro 54
Db 84 GGGAAATTATGAGTAGAG-----GTGGAGCGCGAGCGGATGTCAGAGTCTCT 131
QY 55 ArgValLeuThrGlnAlaSerAsn-----ProValVal----- 65
Db 132 GAATAGTACCATGGGGGAAATGATCCGCTCTGTGTGAAGCCCTTCTCATTCGGA 191
QY 66 -----CysThrGlnProLysSerProSerGlyThrValCys 77
Db 192 TCGCTTTTGGCTTGATGATTGAAATAAGTCTGTGACACAGATGCAGATGCTGTT 251
QY 78 ThrSerLysThrLysLysAlaLeuCysIleThrLeu-----ThrLeuGly 92
Db 252 GTGCACAGATCTCTCTACTCTGCTCCATTTGAAGTTTTCCTCCCAATCATCTGCTGGGG 311
QY 93 ThrPheLeuValGlyAlaAlaLeuAlaGlyLeuLeuTyrLysPheMetGlySerLys 112
Db 312 ATCATTTGATGATTAGACTGGCCATTGCTGGGCATCCACTTC-----GAC 362
QY 113 CysSerAsnSerGlyIleGluCysAspSerSerGlyThrCysIleAsnProSerAsnTyr 132
Db 363 TGCTCAGGGAAG-----TACAGATGTGCTCATCTCTTAAGTGTATCGAGCTGATACTCGA 419
QY 133 CysAspGlyValSerHisCysProGlyGlyGluAspGluAsnArgCysValArgLeuTyr 152

Db 420 TGTGACGGAGTCTCGGATTGCAAGACGGGGAGACGAGTACCGCTGTGTCCGGTGGGT 479
QY 153 GlyProAsnPheIleLeuGlnValTyrSerSerGlnArgLysSerThrPheIleProValCys 172
Db 480 GGTCAAGATGCGCTGCTCAGGTGTTCACAGCT-----GCTTCGTGGAAGACCATGTGC 533
QY 173 GlnAspAspTyrAsnGluAsnTyrGlyArgAlaAlaCysArgAspMetGlyTyrLysAsn 192
Db 534 TCCGATGACTGGAAGGGTCACTAGCAAAATGTTGCTGTGCCCACTGGGTTCCTCA--- 590
QY 193 AsnPheTyrSerSerGlnGlyIleValAspSerSerGlySerThrSerPheMetLysLeu 212
Db 591 AGCTATGTGATTCAGATAAATCAGAGTGCCTGCTGAGGGGGCAGTTCCGGAGAGAG 650
QY 213 AsnThrSerAlaGlyAsnVal-----AspIleTyrLysLysLeuTyrHisSer--- 228
Db 651 TTTGTGTCATTCATCATCTCTTCCAGATGACAGGTGACTGCATTACACCACTCAGTA 710
QY 229 -----AspAlaCysSerSerLysAlaValSerLeuArgCysIleAlaCysGly 245
Db 711 TATGTGAGGGAGGATGTGCTCTGGCCACGTGTGTACCTTGCAGTGCACAGCGCTGTGGT 770
QY 246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGly 265
Db 771 CATGAGGGGCTACAGC-----TCACGCATCGTGGGTGGAACATGCTTGTCTCTCG 824
QY 266 AlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
Db 825 CAGTGGCCCTGGCAGGGCCAGCCTTCAGTTCAGGGCTTACCCTGTGGGGGCTCTGTG 894
QY 286 IleThrProGluTyrIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnPro 305
Db 885 ATCAGCCCTCTGTGATCATCATCTGCTGCACACTGTGTT---TATGACTTGTACCTCCCC 941
QY 306 TrpHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAla--- 324
Db 942 AAGTCATGGACCATCCAGGTGGGTCTAGTT-----TCCCTGTGGCAATATCCAGCCCCA 995
QY 325 GlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLysThrLysAsnAsn 344
Db 996 TCCACACTGTGTGGAGAGATGTGTACACAGCAAGTACAGCAAGAGAGCTGGGCAAT 1055
QY 345 AsplleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProVal 364
Db 1056 GACATCGCCCTTATGAAGCTGGCGGGCCACTCACGTTCAATGAATGATCCAGCCTGTG 1115
QY 365 CysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyrIleSerGlyTyr 384
Db 1116 TGCCTGCCCACTCTGAAGAGAACTTCCCGCATGGAAAGTGTGCTGGACGTGAGATGG 1175
QY 385 GlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuLeu 404
Db 1176 GGGGCCACAGAGGATGAGGTGAGCGCTCCCTGTCTGAACCAACCGCGCCGCTCCCTTTG 1235
QY 405 IleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMet 424
Db 1236 ATTTCCCAAGATCTGCAACACAGGAGCTGTACGGTGGCATCATCTCCCTCCCATG 1295
QY 425 IleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnClyAspSerGlyLysPro 444
Db 1296 CTCGCGGGGCTACCTGACGGGTGGCGTGAACAGCTGCAGGGGACAGCGGGGGGCC 1355
QY 445 LeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSerTyrGlySerGly 464
Db 1356 CTGGTGTCTCAAGAGAGAGGCTGTGGAAGTTAGTGGAGCGACCAAGCTTTGGCATCGC 1415
QY 465 CysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPheThrAspTrpIle 484
Db 1416 TGCCAGAGGTGAACAGCTGGGGGTACACCCGCTGTACCTCTCTTCTTGACTGGATC 1475
QY 485 TyrArgGlnMetArgAlaAsp 491
Db 1476 CACGAGCAGATGGAGAGAGAC 1496


```
RESULT 12
US-09-261-416-1
; Sequence 1, Application US/09261416A
; Patent No. 6291563
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261.416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: 144..1511
; OTHER INFORMATION: CDS
US-09-261-416-1
Alignment Scores:
Pred. No.: 2,02e-74 Length: 2416
Score: 858.00 Matches: 206
Percent Similarity: 52.75% Conservative: 63
Best Local Similarity: 40.39% Mismatches: 186
Query Match: 31.58% Indels: 55
DB: 3 Gaps: 15
US-09-615-285B-2 (1-492) x US-09-261-416-1 (1-2416)
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrVal----- 36
Db 24 AAGCCAGTAACACTGGCGCCCTATCTCTCCGTGGTGCCTACATATTTTGGGACTC 83
Qy 37 -----TyrGluValHisProAlaGlnTyrTyrProSerProValProGlnTyrAlaPro 54
Db 84 GGGNAATATGAGGTAGAG-----GTGGAGGCGGAGCGGATGTCAGAGTCT 131
Qy 55 ArgValLeuThrGlnAlaSerAsn-----ProValVal----- 65
Db 132 GAAATAGTACCACCTGGGGGAAATGATCCGCTCTGTGTTGAAGCCCTTCTCATTCGA 191
Qy 66 -----CysThrGlnProLysSerProSerGlyThrValCys 77
Db 192 TCGCTTTTGGCCTTGATGATTGAAATAGTCTCTGTCACACAGATGCTGTT 251
Qy 78 ThrSerLysThrLysLysAlaLeuCysAlaThrLeu-----ThrLeuGly 92
Db 252 GCTGCACAGATCCTGCTCACTGCTGCCATTTGAAGTTTTCCTCCCAATCATGCTCATTTGGG 311
Qy 93 ThrPheLeuValGlyAlaAlaLeuAlaGlyLeuLeuTrpLysPheMetGlySerLys 112
Db 312 ATCATTCGATGATATTAGACATGCGCCATTTGGTCTGGGCATTCACCTTC-----GAC 362
Qy 113 CysSerAsnSerGlyIleGluCysAspSerSerGlyThrCysIleAsnProSerAsnTyr 132
Db 363 TGCTCAGGAG--TACAGATGTCGCTCATCTTTAAGTGTATCGAGCTGATAACTCGA 419
Qy 133 CysAspGlyValSerHisCysProGlyGlyGluAspGluAsnArgCysValArgLeuTyr 152
Db 420 TGTGACGGAGTCTCGGATTGCAAGACGGGGAGGACGAGTACCGCTGTGTCCGGGTGGGT 479
Qy 153 GlyProAsnPheLeuLeuGlnValTyrSerSerGlnArgLysSerTrpHisProValCys 172
Db 480 GGTGAGAAATGCGGTCTCCAGGTGTTCCACAGCT-----CCTTCGTGGAGACCATGTGC 533
Qy 173 GlnAspAspTrpAsnGluAsnTyrGlyArgAlaAlaCysArgAspMetGlyTyrLysAsn 192
Db 534 TCCGATGACTGGAAGGGTCACTACGCAATGTTGCTGTGCTCCCAACTGGGTTCCTCCCA--- 590
```

```
Qy 193 AsnPheTyrSerSerGlnGlyIleValAspAspSerGlySerThrSerPheMetLysLeu 212
Db 591 AGCTATGTGAGTTCAGATAACCTCAGAGTACGCTCGTGGAGGGCAGTTCGCGGAGGAG 650
Qy 213 AsnThrSerAlaGlyAsnVal-----AspIleTyrLysLysLeuTyrHiser--- 228
Db 651 TTTGTGTCATCGATCAGTACCTCTTCCAGATGACAGGTGACTGATACCATCAGTACGTA 710
Qy 229 -----AspAlaCysSerSerLysAlaValValSerLeuArgCysIleAlaCysGly 245
Db 711 TATGTGAGGAGGAGGATGTGCTCTGGCCACGTGTTACTTGCAGTGCACAGCCTGTGTGT 770
Qy 246 ValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGluSerAlaLeuProGly 265
Db 771 CATAGAAGGGGCTACAGC-----TCAGCATCTGGGTGGAAACATGCTCTTCTCTCG 824
Qy 266 AlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisValCysGlyGlySerIle 285
Db 825 CAGTGGCCCTGGCAGGCCAGCTTTCAGTTCAGGGCTACCACTGTGGGGGCTCTGTC 884
Qy 286 IleThrProGluTrpIleValThrAlaAlaHisCysValGluLysProLeuAsnAsnPro 305
Db 885 ATCAGCCCTCTGTGGATCATCTGCTGCACACCTGTGT-----TATGACTTGTACCTCC 941
Qy 306 TrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMetPheTyrGlyAla--- 324
Db 942 AAGTCATGACCATCCAGGTGGTCTAGTT-----TCCCTGTTGGACAATCCAGCCCA 995
Qy 325 GlyTyrGlnValGluLysValIleSerHisProLysTyrAspSerLysThrLysAsnAsn 344
Db 996 TCCACATTTGGTGGAGAGATTGTTTACCACAGCAAGTACAAAGCCCAAGAGCTGGGCAAT 1055
Qy 345 AspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeuValLysProVal 364
Db 1056 GACATCGCCCTTATGAGTGGCGGGCCACTCAGCTTCATGAAATGATCCAGCTGTG 1115
Qy 365 CysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrpIleSerGlyTrp 384
Db 1116 TGCCTGCCCAACTCTGAAGAGAACTTCCCCCATGGAAGTGTCTGACGCTCAGGATGG 1175
Qy 385 GlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAlaLysValLeuLeu 404
Db 1176 GGGGCCACAGAGAGTGGAGGTGAGCTCCCTCTCTGACACCGCGCGCTCCCTTTG 1235
Qy 405 IleGlu-ThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuLeuThrProAlaMe 424
Db 1236 ATTTCCAAAGATGCTCAACACAGGAGCTGTACGTTGGCATCATCTCCCTCCAT 1295
Qy 424 IleCysAlaGlyPheLeuGlnGlyAsnVal--AspSerCysGlnGlyAspSerGlyGly 443
Db 1296 GCTCTGCGCGGCTACCTGACGGGTGGAGTGGAACTCCAGGGGAGCAGCGGGGG 1355
Qy 444 ProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSerTrpGlySer 463
Db 1356 CCCCTGGTGTCTCAAGAGAGGAGCTGTGGAAGTTAGTGGAGCGACAGCTTTGGCATC 1415
Qy 464 GlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPheThrAspTrp 483
Db 1416 GGCTGCGCAGACGTGAACAGCCTGGGTGTACACCGGTGTACCTCCTCTCTGACTGG 1475
Qy 484 IleTyrArgGlnMetArgAlaAsp 491
Db 1476 ATCCAGCAGCATGAGAGAGAC 1499
```

RESULT 13

```
US-09-656-002-1
; Sequence 1, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
```


; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-69108/DJB/JDD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US/09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US/09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-656-002-1

Alignment Scores:
Pred. No.: 9,71e-58 Length: 2079
Score: 687.50 Matches: 171
Percent Similarity: 48.86% Conservative: 65
Best Local Similarity: 35.40% Mismatches: 170
Query Match: 25.30% Indels: 77
DB: 4 Gaps: 18

US-09-615-285B-2 (1-492) x US-09-656-002-1 (1-2079)

Qy 41 ProAlaGlnTyrProSerProValProGlnTyrAlaProArgValLeuThr----- 58
Db 203 CCAGCATGGTAC-----AGGATCCTCACAGTGATC 232
Qy 59 -----GlnAlaSerAsnProValValCysThrGlnProLysSerProSer 73
Db 233 AACCTCTGAACAGCCTCGATGTAACCC-----TGGCAAAACCCCGTATCCCCA--- 283
Qy 74 GlyThrValCysThrSerLysThrLysAlaLeuCysIleThrLeuThrLeuGlyThr 93
Db 284 TGGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCCTGGCGAGT 343
Qy 94 PheLeuValGlyAlaAlaLeuAlaLeuAlaGlyLeuLeuTrpLysPheMetGlySerLysCys 113
Db 344 ATCATCTGTGGTGTCTCTCATCAGGTGATCTTGGATAAATACTAC----- 391
Qy 114 SerAsnSerGlyIleGluCysAspSerSerGlyThrCysIleAsnProSerAsnTrpCys 133
Db 392 -----TTCCCTCTGGGGCAGCCTCTCCACITCATCCCGAGGAAGCAGCTGTGT 439
Qy 134 AspGlyValSerHisCysProGlyGlyGluAspGluAsnArgCysValArgLeuThr--- 152
Db 440 GACGAGAGCTGGACTGTCTCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCC 499
Qy 153 ---GlyPro-----AsnPheIleLeuGlnValTyrSer 162
Db 500 GAAGGCTCTGAGTGGCAGTCGGCTCTCCAGGACCGATCCACACTGCGGTCTGGAC 559
Qy 163 SerGlnArgLysSerTrpHisProValCysGlnAspTrpAsnGluAsnTyrGlyArg 182
Db 560 TCGGCCACAGGAAGTGTCTCTGCTGTTTTCACAACTTCACAGAAAGCTCTCGCTGAG 619
Qy 183 AlaAlaCysArgAspMetGlyTyr-----LysAsnAsnPheTyrSerSerGln----- 198
Db 620 ACAGCTCTAGCAGATGGCTTACAGCAGCAACCCACTTCAGAGCTGTGGAGATTGGC 679
Qy 199 -----GlyIleValAspAspSerGlySerThrSerPheMetLysLeuAsn 213
Db 680 CCAGACAGGATCTGGATGTTGTTGAATATCACAGAAACACAGCCAGGAGCTTCGCATCGG 739
Qy 214 ThrSerAlaGlyAsnValAspIleTyrLysLysLeuTyrHisSerAspAlaCysSerSer 233
Db 740 AACTCAAGTGG-----CCTGTCTCTCA 763
Qy 234 LysAlaValSerLeuArgCysIleAlaCysGlyValAsnLeuAsnSerSerArgGln 253

Db 764 GGCTCCCTGCTGCTCCCTGCACCTGTCTGCTGTGGAGAGCCTG-----AAGACC 814
Qy 254 SerArgIleValGlyGlyGluSerAlaLeuProGlyAlaTrpProTrpGlnValSerLeu 273
Db 815 CCCGCTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGCCCTGGCAGTGCAGATC 874
Qy 274 HisValGlnAsnValHisValCysGlyGlySerIleThrProGluTrpIleValThr 293
Db 875 CAGTACGACAAACAGCAGCTCTGTGAGGAGGAGCATCTGGACCCCACTGGTCTTCACG 934
Qy 294 AlaAlaHisCysValGluLysProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGly 313
Db 935 GCAGCCCACTGCTTCAGGAACAT---ACCGATGTGTTCACTGGAGGTGGCGGAGC 991
Qy 314 IleLeuArg---GlnSerPheMetPheTyrGlyAlaGlyTyrGlnValGluLysValIle 332
Db 992 TCAGACAAACTGGCAGCTTC-----CCATCCCTGGCTGTGCGCCAGATCATC 1039
Qy 333 -----SerHisProAsnTyrAspSerLysThrLysAsnAsnAspIleAlaLeuMet 349
Db 1040 ATCATTTGAATTCAACCCCATGTAC-----CCCAAGACAAATGACATGCCCTCATG 1090
Qy 350 LysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnPro 369
Db 1091 AAGCTGCACTTCCCACTCACTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTT 1150
Qy 370 GlyMetMetLeuGlnProGluGlnLysTrpIleSerGlyTrpGlyAlaThrGluGlu 389
Db 1151 GATGAGGACTCACTCCAGCCCACTCTGGATCATTTGGATGGGCTTTACGAAGCAG 1210
Qy 390 Lys---GlyLysThrSerGluValLeuAenAlaAlaLysValLeuLeuIleGluThrGln 408
Db 1211 AATGGAGGGAAGATGCTGACATCTGCTGAGCGCTCAGTCCAGTCTTTCAGACACA 1270
Qy 409 ArgCysAsnSerArgTyrValTyrAspAsnLeuIleThrProAlaMetIleCysAlaGly 428
Db 1271 CGGTGCAATGCAAGCAGATCGTACCAGGGGAGGATCCAGAGAGATGATGTGCGAGC 1330
Qy 429 PheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGlyProLeuValThrSer 448
Db 1331 ATCCCGAAGGGGTGTGGACACCTGCGAGGTGACAGTGGTGGGCCCTGATGTACCAA 1390
Qy 449 LysAsnAsnIleTrpTrpLeuIleGlyAspThrSerTrpGlySerGlyCysAlaLysAla 468
Db 1391 TCTGACAG---TGGCATGTGGTGGCATCGTTAGTGGGGCTATGGCTGGCGGGCCCG 1447
Qy 469 TyrArgProGlyValTyrGlyAsnValMetValPheThrAspTrpIleTyrArgGlnMet 488
Db 1448 AGCACCCAGGAGTATACCAAGGTCTCAGGCTATCTCACTGGATCTACATGTCTGG 1507
Qy 489 ArgAlaAsp 491
Db 1508 AAGGCTGAG 1516

RESULT 14
US-09-651-588-5
; Sequence 5, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; FILE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JDD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US/09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US/09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 2079
TYPE: DNA
ORGANISM: Homo sapiens
US-09-851-588-5

Alignment Scores:

Pred. No.: 9,71e-58 Length: 2079
Score: 687.50 Matches: 171
Percent Similarity: 48.86% Conservative: 65
Best Local Similarity: 35.40% Mismatches: 170
Query Match: 25.30% Indels: 77
DB: 4 Gaps: 18

US-09-615-285B-2 (1-492) x US-09-851-588-5 (1-2079)

Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThr----- 58
Db 203 CCAGCATGGTAC-----AGGATCCTGCAGCATGATC 232
Qy 59 -----GlnAlaSerAsnProValValCysThrGlnProLysSerProSer 73
Db 233 AACCTCGAACAGCCTCGATGTCAAACCCC-----TSCGCAACCCCGTATCCCCA--- 283
Qy 74 GlyThrValCysThrSerLysThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThr 93
Db 284 TGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCATTCTAGACCTTGGCGAGT 343
Qy 94 PheLeuValGlyAlaAlaLeuAlaAlaGlyLeuLeuThrLysPheMetGlySerLysCys 113
Db 344 ATCATCATGTGGTGTCCCTCATCAAGGTGATTCGATAAATACTAC----- 391
Qy 114 SerAsnSerGlyIleGluCysAspSerSerGlyThrCysIleAsnProSerAsnTrpCys 133
Db 392 -----TTCCCTCTGGGGCGCCCTCCCATCTTCATCCCGAGGAGCAGCTGTGT 439
Qy 134 AspGlyValSerHisCysProGlyGlyLysAspGlnAsnArgCysValArgLeuTyr--- 152
Db 440 CACGAGAGCTGGAGCTGTCCCTTGGGGAGGACGAGGAGCAGTGTGTCAAGAGCTTCCCC 499
Qy 153 ---GlyPro-----AsnPheLeuLeuGlnValTyrSer 162
Db 500 GAAGGCGCTGCAGTGGCGAGTCCGCTCTCCAGAGCCGATCCACACTGCAGGTGTGGAC 559
Qy 163 SerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyrGlyArg 182
Db 560 TCGGCCACAGGAACTGTTCTCTGCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAG 619
Qy 183 AlaAlaCysArgAspMetGlyTyr-----LysAsnAsnPheTyrSerSerGln----- 198
Db 620 ACAGCCTGTAGGCAGATGGGTACAGCGAAACCCACTTCAGAGCTGTGGAGATTGGC 679
Qy 199 -----GlyIleValAspAspSerGlySerThrSerPheMetLysLeuAsn 213
Db 680 CCAGACACAGGATCTGGATGTTGTAATCATCAGAAACAGCCAGGAGCTTCGATCGG 739
Qy 214 ThrSerAlaGlyAsnValAspIleTyrLysLysLeuTyrHisSerAspAlaCysSerSer 233
Db 740 AACTCAAGTGGG-----CCCTGTCTCTCA 763
Qy 234 LysAlaValValSerLeuArgCysIleAlaCysGlyValAsnLeuAsnSerSerArgGln 253
Db 764 GGCTCCCTGGTCTCCCTGCAGTCTGCTGCTGCTGGAGAGCCTG-----AAGACC 814
Qy 254 SerArgIleValGlyGlySerAlaLeuProGlyAlaTrpProTrpGlnValSerLeu 273
Db 815 CCCCGTGTGGTGGGTGGGAGGAGCCCTCTGTGGATTCTTGGCCCTTGGCAGGTGAGCATC 874
Qy 274 HisValGlnAsnValHisValCysGlyGlySerIleThrProGluTrpIleValThr 293
Db 875 CAGTACGACAAACACACACGATCTGTGGAGGAGCATCTCTGGACCCCTGAGTCTGCTCAGC 934

Qy 294 AlaAlaHisCysValGluLysProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGly 313
Db 935 GCAGCCCACTGCTTCAGGAACAT---ACCAGTGTCTCACTGGAAGTCCGGGCGAGGC 991
Qy 314 IleLeuArg---GlnSerPheMetPheTyrGlyAlaGlyTyrGlnValGluLysValIle 332
Db 992 TCAGACAAACTGGGCGAGCTTC-----CCATCCCTGGCTGTGGCCCAAGATCATC 1039
Qy 333 -----SerHisProAsnTyrAspSerLysThrLysAsnAsnAspIleAlaLeuMet 349
Db 1040 ATCATTTGAATTCACCCCATGTAC-----CCCAAGACATGACATCGCCTCATG 1090
Qy 350 LysLeuGlnLysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnPro 369
Db 1091 AAGCTGCAGTTCCTCCACTCTCTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTT 1150
Qy 370 GlyMetMetLeuGlnProGluLeuLysCysTrpIleSerGlyTyrGlyValThrGluGlu 389
Db 1151 GATGAGGAGCTACTCTCCAGCCACCCCATCTGTGATCATTTGGGCTTTACGAAGCAG 1210
Qy 390 Lys---GlyLysThrSerGluValLeuAsnAlaLysValLeuLeuIleGluThrGln 408
Db 1211 AATGGAGGGAAGATGCTGTGACATACTGTCTGAGGCGTCAGTCCAGTCAATTGACACACA 1270
Qy 409 ArgCysAsnSerArgTyrValTyrAspAsnLeuLeuThrProAlaMetIleCysAlaGly 428
Db 1271 CGGTGCAATGCGACAGCATCGCTACCGGGGAAGTCACCGAGAGATGATGTGCGAGGC 1330
Qy 429 PheLeuGlnGlyAsnValAspSerCysGlnGlyAspSerGlyGlyProLeuValThrSer 448
Db 1331 ATCCCGGAAGGGGTGTGGACACCTCCAGGCTGACAGTGTGGGCCCTGATGTACCAA 1390
Qy 449 LysAsnAsnIleTrpTrpLeuIleGlyAspThrSerTrpGlySerGlyCysAlaLysAla 468
Db 1391 TCTGACCAAG---TGGCATGTGTGGGCATCGTTAGCTGGGCTATGCTGCGGGGCGCG 1447
Qy 469 TyrArgProGlyValTyrGlyAsnValMetValPheThrAspTrpIleTyrArgGlnMet 488
Db 1448 AGCACCCAGGAGTATACACCAAGCTCTCAGCCTATCTCACTGGATCTACAATGTCTGG 1507
Qy 489 ArgAlaAsp 491
Db 1508 AAGGCTGAG 1516

RESULT 15

US-09-008-271A-18

Sequence 18, Application US/09008271A

Patent No. 6203979

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Hillman, Jennifer L.

Yue, Henry

Guegler, Karl J.

Corley, Neil C.

Tang, Tom Y.

Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESS: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/008,271A

FILING DATE: 16-Jan-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PP-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT13
CLONE: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-008-271A-18

Alignment Scores:
Pred. No.: 1.4e-56 Length: 2038
Score: 676.50 Matches: 166
Percent Similarity: 50.22% Conservative: 65
Best Local Similarity: 36.09% Mismatches: 161
Query Match: 24.90% Indels: 69
Gaps: 17

US-09-615-285B-2 (1-492) x US-09-008-271A-18 (1-2038)

QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 240 TCRAACCC-----TCCGCAACCCCGTATCCCATG-CAGACCTTC-----280
QY 81 ThrLysLysala-----LeuCysileThrLeuThrLeuGlyThrPheLeuVal 96
DB 281 ---AGAAAGTGGGGATCCCATCATCATGACACTACTGAGCCTGCGAGTATCATCAT 337
QY 97 GlyAlaAlaLeuAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSer 116
DB 338 GTGGTTCTCTCATCAAGGTGATCTGGATTAATCTAC-----376
QY 117 GlyIleGluCysAspSerSerGlyThrCysileAsnProSerAsnTrpCysAspGlyVal 136
DB 377 ---TTCTCTTGGGGAGCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGAGAG 433
QY 137 SerHisCysProGlyGlyGluAspGluAsnArgCysValargLeuTyr-----GlyPro 154
DB 434 CTGGACTGTCCCTTGGGGGAGGAGCAGGAGCAGTGTGTCAAGAGCTTCCCGGAAGGGCCT 493
QY 155 -----AsnPheileLeuGlnValTyrSerSerGlnArg 165
DB 494 GCAGTGGCAGTCCGCCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACA 553
QY 166 LysSerTrpHisProValCysGlnAspTrpAsnGluAsnTyrGlyArgAlaAlaCys 185
DB 554 GGGAACTGGTCTCTGCTGCTGTTTCGACAACTTACAGAGAGCTCTCGCTGAGACAGCCTGT 613
QY 186 ArgAspMetGlyTyr-----LysAsnAsnPheTyrSerSerGln-----198
DB 614 AGGCAGATGGGCTACAGCAAAACCCACTTTAGAGCTGTGGAGATTGGCCAGACCAG 673
QY 199 -----GlyIleValAspSerGlySerThrSerPheMetLysLeuAsnThrSerAla 216
DB 674 GATCTGGATGTTGTTGAATCACAGAAACAGCCAGGAGCTTCGATGCGGAACCTCAAGT 733
QY 217 GlyAsnValAspIleTyrLysLeuTyrHisSerAspAlaCysSerSerLysAlaVal 236
DB 734 GGG-----CCCTGTCTCTCAGGCTCCCTG 757
QY 237 ValSerLeuArgCysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIle 256

758 GTCTCCCTGCACCTGTCTTGCCTGTGGGAGAGCCTG-----AAGACCCCGCTGTG 808
QY 257 ValGlyGlyGluSerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGln 276
DB 809 GTGGTGGGAGAGGCGCTCTGTGTGATTTCTTGGCTTGGCAGGTACAGCTACGAC 868
QY 277 AsnValHisValCysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHis 296
DB 869 AAACGACGAGCTGTGGAGGAGCATCTGTGACCCCACTGGGTCTCAGGAGCCAC 928
QY 297 CysValGluLysProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArg 316
DB 929 TGCTTCAGAAACAT---ACGATGTGTCACTGGAAGGTGCGGCGAGGCTCAGACAAA 985
QY 317 ---GlnSerPheMetPheTyrGlyAlaGlyTyrGlnValGluLysValIle-----332
DB 986 CTGGGCAGCTTC-----CCATCCCTGGCTGTGGCCAAGATCATCATATTGAA 1033
QY 333 SerHisProAsnTyrAspSerLysThrLysAsnAsnAspIleAlaLeuMetLysLeuGln 352
DB 1034 TTCRAACCCCATGTAC-----CCCAAGACAATGACATCGCCCTCATGAAGCTGAG 1084
QY 353 LysProLeuThrPheAsnAspLeuValLysProValCysLeuProAsnProGlyMetMet 372
DB 1085 TTCCCACTCATTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTTCTTGTATGAGAG 1144
QY 373 LeuGlnProGluGlnLeuCysTrpIleSerGlyTrpGlyAlaThrGluGluLys---Gly 391
DB 1145 CTCACTCCAGCCACCCCACTCTGATCATTTGATGGGCTTTACGAGCAGAATGAGGG 1204
QY 392 LysThrSerGluValLeuAsnAlaAlaLysValLeuLeuIleGluThrGlnArgCysAsn 411
DB 1205 AAGATGTCTGACATACTCTGCAGCGCTCAGTCCAGGTTCATTGACAGCACCGGTGCAAT 1264
QY 412 SerArgTyrValTyrAspAsnLeuIleThrProAlaMetIleCysAlaGlyPheLeuGln 431
DB 1265 GCAGACGATGGTACCCAGGGGAAGTCCACCAGAGATGATGTGTGAGGCATCCCGGAA 1324
QY 432 GlyAsnValAspSerCysGlnGlyAspSerGlyGlyProLeuValThrSerLysAsnAsn 451
DB 1325 GGGGTGTGGACACCTGCCAGGTGACAGTGTGGGCCCTGTGTACCAATCTGACCAG 1384
QY 452 IleTrpTrpLeuIleGlyAspThrSerTrpGlySerGlyCysAlaLysAlaTyrArgPro 471
DB 1385 ---TGGCATGTGGTGGGCATCGTTAGCTGGGGCTATGGCTCGGGGGCCCGACACCCCA 1441
QY 472 GlyValTyrGlyAsnValMetValPheThrAspTrpIleTyrArgGlnMetArgAlaAsp 491
DB 1442 GGAGTATACCAAGGTCTCAGCCTATCTCACTGATCTACTAATGTCTGGAAGCTGAG 1501

Search completed: June 1, 2004, 17:42:16
Job time : 159 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: June 1, 2004; 14:38:29 / Search time 629 seconds
(without alignments)
3322.915 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 2717
Sequence: 1 MALNSGSPAPGYYENHGY.....VGNVMVFTDWYQMRADG 492

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODE=frame+ p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool/US09615285/runat 01062004 132132 19940/app query.fasta_1.647
-DB=N_Geneseq 257at04 -OPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09615285 @CGN 1 1 470 @runat 01062004 132132 19940 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04: *
1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002s: *
7: Geneseqn2003as: *
8: Geneseqn2003bs: *
9: Geneseqn2003cs: *
10: Geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2717	100.0	1479	3	Aaz87786 Human tum
2	2717	100.0	1738	3	Aaz29636 Human 20P
3	2717	100.0	1738	6	Aad28778 Human 20P
4	2712	99.8	2486	8	Adal0904 Human CDN
5	2712	99.8	2486	8	Acho03998 Human CDN
6	2711	99.8	3226	9	Adb75590 Prostate
7	2709	99.7	1476	4	Aah93944 P1000C op
8	2709	99.7	1476	4	Aas64180 Human pro

9	2709	99.7	1476	5	ACA59988	ACA59988 Prostate
10	2709	99.7	1476	6	ABL95551	ABL95551 Human P10
11	2709	99.7	1476	7	ACC95715	ACC95715 Prostate
12	2709	99.7	1476	9	ADBI4381	ADBI4381 Human pro
13	2709	99.7	1479	4	Aah93943	Aah93943 P1000C op
14	2709	99.7	1479	4	AAS64179	AAS64179 Human pro
15	2709	99.7	1479	5	ACA59987	ACA59987 Prostate
16	2709	99.7	1479	6	ABL95550	ABL95550 Human P10
17	2709	99.7	1479	7	ACC95714	ACC95714 Prostate
18	2709	99.7	1479	9	ADBI4380	ADBI4380 Human pro
19	2709	99.7	3245	3	AAA08803	AAA08803 Androgen-
20	2709	99.7	3245	4	Aah93942	Aah93942 P1000C fu
21	2709	99.7	3245	4	AAS64178	AAS64178 Human pro
22	2709	99.7	3245	5	ACA59986	ACA59986 Prostate
23	2709	99.7	3245	6	ABL95549	ABL95549 Human P10
24	2709	99.7	3245	7	ACC95713	ACC95713 Prostate
25	2709	99.7	3245	9	ADBI4379	ADBI4379 Human pro
26	2709	99.7	3483	5	ABV21748	ABV21748 Human pro
27	2709	99.7	3483	5	ABV23318	ABV23318 Human pro
28	2709	99.7	3483	5	ABV24651	ABV24651 Human pro
29	2709	99.7	3483	5	ABV29119	ABV29119 Human pro
30	2709	99.7	3483	5	ABV27570	ABV27570 Human pro
31	2709	99.7	3483	5	ABV21767	ABV21767 Human pro
32	2709	99.7	3483	5	ABV29165	ABV29165 Human pro
33	2709	99.7	3483	5	ABV27589	ABV27589 Human pro
34	2704	99.5	2473	3	Aaz95005	Aaz95005 Cancer sp
35	2696	99.2	2479	3	Aaz87813	Aaz87813 Human tum
36	2696	99.2	2479	3	AAH93928	AAH93928 Human tra
37	2696	99.2	2479	4	AAS64164	AAS64164 Human CDN
38	2696	99.2	2479	4	AAD13168	AAD13168 Human ser
39	2696	99.2	2479	5	ACA59972	ACA59972 Prostate
40	2696	99.2	2479	6	ABL95535	ABL95535 Human tra
41	2696	99.2	2479	6	ABK92201	ABK92201 Prostate
42	2696	99.2	2479	6	AAD28779	AAD28779 Human TMP
43	2696	99.2	2479	7	ACC51058	ACC51058 Human bla
45	2696	99.2	2479	7	ACC95699	ACC95699 Human tra

ALIGNMENTS

RESULT 1
AAZ87786
ID AAZ87786 standard; DNA; 1479 BP.
XX
AC AAZ87786;
DT 05-JUL-2001 (revised)
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMPRSS2 encoding DNA.
XX
KW Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
KW gene therapy; protein therapy; ds.
XX
OS Homo sapiens.
XX
FH Key
CDS Location/Qualifiers
FT 1..1479
FT /tag= a
FT /product= "TMPRSS2"
FT 478
FT /tag= b
FT /note= "this base can be G or A with G being the more
FT common allele. The codon will change from Val to Met."
FT 625
FT /tag= c
FT /note= "this base can be T or A with T being the more
FT common allele. The codon will change from Phe to Ile"
FT 724
FT conflict
FT /tag= d
FT /note= "listed as T in GenBank Accn No: U75329"
FT 768
FT allele

FT FT /tag= e
 FT FT /note= "this base can be C or T with C being the more
 FT FT common allele. This is a silent polymorphism"
 FT FT 777
 FT FT allele
 FT FT /tag= f
 FT FT /note= "this base can be C or T with C being the more
 FT FT common allele. The codon is unaffected with both alleles
 FT FT encoding Gly"
 FT FT 834
 FT FT allele
 FT FT /tag= g
 FT FT /note= "this base can be C or T with C being the more
 FT FT common allele. This is a silent polymorphism"
 FT FT 985
 FT FT conflict
 FT FT /tag= h
 FT FT /note= "listed as C in GenBank Accn No: U75329"
 FT FT 1347
 FT FT conflict
 FT FT /tag= i
 FT FT /note= "listed as C in GenBank Accn No: U75329"
 FT FT 1466
 FT FT conflict
 FT FT /tag= j
 FT FT /note= "listed as A in GenBank Accn No: U75329"
 FT FT 1471
 FT FT conflict
 FT FT /tag= k
 FT FT /note= "listed as A in GenBank Accn No: U75329"
 FT FT 1471
 XX W020000605-A1.
 PN 06-JAN-2000.
 PD 29-JUN-1999; 99WO-US014622.
 XX 29-JUN-1999; 98US-0091044P.
 PR (MYRI-) MYRIAD GENETICS INC.
 PA Wong AKC, Tavtigian SV, Teng DHF;
 PI WPI; 2000-170914/15.
 DR P-PSDB; AAY7726.
 XX Novel tumor suppressor TMRSS2 used for the diagnosis and prognosis of
 PT human cancer.
 XX Claim 22; Page 77-79; 89pp; English.
 PS The invention provides a new tumour suppressor gene, designated TMRSS2.
 CC The TMRSS2 polynucleotides and polypeptides can be used in methods for
 CC diagnosing and prognosing predisposition to cancer in humans. The
 CC polypeptides may also be used in assays to screen for compounds with anti
 CC -cancer or therapeutic properties. The polypeptides are also useful for
 CC rational drug design. The TMRSS2 polynucleotides and polypeptides may be
 CC used for gene therapy and protein therapy. The present sequence
 CC represents the DNA encoding the TMRSS2 polypeptide. (N.B. Revised entry
 CC submitted to correct residues 241-245 of the sequence.)
 XX SQ Sequence 1479 BP; 363 A; 407 C; 402 G; 307 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 7,46e-206 Length: 1479
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0
 US-09-615-285B-2 (1-492) x AAZ87786 (1-1479)
 QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 DB 1 ATGGCTTTGAACTCAGGGTCACCAACAGCTATTGGACCTTACTATGAAACCATGATAC 60
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40

Db 61 CAACCGGAAAACCCCCTATCCCGCAGCCACCTGTGGTCCCCCACTGTCTTACGAGGTGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CCGGCTCAGTACTACCCGTCGCCCGCCAGTACGCCCGGAGGTCTCTGACGAGGT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCCAAACCCCGTGTGTGACGAGCCCAATCCCATCCGACAGTGTGCACCTCAAG 240
 QY 81 ThrLysLeuAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 Db 241 ACTAAGAAAGCACTGTGCATCCTTACCTGGGACCTTCTCTGTGGAGGTGGCTG 300
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 301 GCGCTGGCCTACTCTGGAAAGTTTCATGGCAGCAAGTGTCTCAACTCTGGGATAGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 361 GACTCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGGTGCACACTGCCCC 420
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 421 GCGGGGAGGACGAGAAATCGGTGTGTTCGCCCTCTACGACCAAACTTCTCTTCAAGTG 480
 QY 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
 Db 481 TACTCATCTCAGAGGAGTCTTGGCACCTGTGTGCCAAGACGACTGGAGAACTAC 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 541 GCGCGGCGCCTGCGAGGACATGGGCTATAAGATAATTTTACTCTTACCAAGAAATA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAACTGNAACACAGTCCCGGCAATGTGAT 660
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 661 ATCTATAAAAAAAGTGTACCACTGATGCTGTCTTCTTCAAAAGCAGTGTGTTCTTACGC 720
 QY 241 CysIleAlaCysGlyValLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 721 TGTATAGCTTGGCGGTCACTTGAATCAAGCCGCGACAGAGGATCGTGGCGCGGAG 780
 QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 781 AGCGCGCTCCCGGGGCTGGCCCTGGCAGGTGAGCTGCGCTCCAGAACGTCACGTCG 840
 QY 281 CysGlyGlySerIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
 Db 841 TCGGAGGCTTCATCATCACCCCGAGTGTGATCGTGCAGCCGCCCTGCGTGGAAAAA 900
 QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 901 CCTCTTAACAATCCATGGCATTTGGACGGCATTTTGGGGGATTTTGAGACAATCTTTCATG 960
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 961 TTCTATGGAGCCGGATACCAAGTAGAAAAAGTATTTTCATCAATAATTATGACTCAAG 1020
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1021 ACCAAGNACAATGACATTTGGCTGATGAAGCTGCAGAGGCTCTGACTTTTCAACGACCTA 1080
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 Db 1081 GTGAAACCAAGTGTGTCTGCCCAACCCAGGCGATGCTGTGCAGCAGAACAGCTCTGTCTG 1140
 QY 381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 Db 1141 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAAAGACCTCAGAGTGTGTGACGCTGCC 1200

CC	20PFI2/TMPRSS2 gene (also designated 20PFI2-GTCL1, as deposited with ATCC accession number 207097). Anti-20PFI2/TMPRSS2 antibodies may be used as therapeutic agent for prostate and colon cancers, to image prostate cancer cells and prostate tumours, to identify ligands and cellular constituents that bind to a 20PFI2/TMPRSS2 gene product and for use as cancer vaccines
XX	
SQ	Sequence 1738 BP; 433 A; 459 C; 473 G; 373 T; 0 U; 0 Other;
Alignment Scores:	
Pred. No.:	9,146-206 Length: 1738
Score:	2717.00 Matches: 492
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	3 Gaps: 0
US-09-615-285B-2 (1-492) x AAZ29636 (1-1738)	
QY	1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB	112 ATGGCTTTGAACTCAGGTGCACCACGACTATGGACCTATTATGAAAAACCATGATAC 171
QY	21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB	172 CAACCGGAAAACCCCTATCCGCACAGGCCACTGTGGTCCCCACTGTCTACGAGGTGCAT 231
QY	41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB	232 CGGGCTCAGTACTACCGGTCCCCTGGCCAGTAGCACGCCCGGAGGGTCTCTGACGCAAGCT 291
QY	61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB	292 TCACAACCGCTGCTGCAGCGAGCCAAATCCCCATCCGGGACAGTGTCACCTCAAAG 351
QY	81 ThrLysLysAlaLeuCysalleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB	352 ACTAAGAAGCACTGTGCATCACCTTGACCTGGGAGACCTTCTCGTGGAGCTGCGCTG 411
QY	101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB	412 GCCGTGGCTACTCTGGAGTTCTATGGCAGCAAGTGCTCCAACCTCTGGATAGAGTGC 471
QY	121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB	472 GACTCTCTCAGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGGTGCACACTGCCCC 531
QY	141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB	532 GCGCGGGAGACGAGAACTGGTGTGTTTCGCTCTACGGACCAAACTTCATCTTCAGGTG 591
QY	161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
DB	592 TACTCATCTCAGAGAAGTCTCTGGCACCCCTGTGTGCCAAGCACTGGAACACGAGAACTAC 651
QY	181 GlyArgAlaAlaCysArgAspMetGlyTyrlLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB	652 GGCGCGGGCGGCTTCAGGAGCATANGGCTATAGAATAATTTTTTCTCTAGCCAGGANTTA 711
QY	201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB	712 GTGGATGACAGCGGATCCACCAGCTTTATGAACCTGAACACAAGTGC CGCAATGTGCGAT 771
QY	221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB	772 ATCTATAAAAACCTGACACACAGTATGCCTGTTCTTCAAAAGCAGTGGTTCCTTACGC 831
QY	241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgileValGlyGlyGlu 260
DB	832 TGATAGCTCGGGGTCAACTTGACITCAAGCCCGCAGAGCAGGATTGTGGCGCGCGAG 891
QY	261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280

Db 892 AGCGCGCTCCGGGGGCTGGCTGGCAGGTGACGCTCCAGAACGCTCCACGCTG 951
Qy 281 CysGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
Db 952 TGGGAGGCTCCATCATCACCCCGAGTGGATCGTGACAGCCGCCCTGCGGAAAA 1011
Qy 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1012 CCTCTTAACCAATCATGCGCATGGACGGCAATTTGGCGGATTTTGAGACAATCTTTCA 1071
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1072 TTTATGAGCGCGGATACCAAGTAGAAAAAGTGAATTTCTCATCAAAATATGACTCCA 1131
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1132 ACCAAGAACCAATGACATTTGGCTGATGAAGCTGCAGAGGCTCTGACTTTCAACGACCTA 1191
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnIleuCysTyr 380
Db 1192 GTGAACCAAGTGTCTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGCTG 1251
Qy 381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1252 ATTTCCGGTGGGGGGCCACGAGGAGAAAGGAGACCTCAGAGTGTGTGAACGTGCC 1311
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1312 AAGGTGCTTCTCATGTGACACAGATGCAACAGCAGATGCTGTATGACAACTGATC 1371
Qy 421 ThrProAlaMetIleCysAlaGlyPheIleuGlnCysValAsnValAspSerCysGlnGlyAsp 440
Db 1372 ACACAGCCATGATCTGTGGCGGCTCTCTGCAGGGGAACTGCGATTTCTCCAGGGTGAC 1431
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
Db 1432 AGTGGAGGGCTCTGGTCACTTCGAAGAACATATCTGGTGGCTGATAGGGGATACAAGC 1491
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1492 TGGGGTCTGCTGCTGGCCAAAGCTTACAGACCCAGAGTGTACGGGAATGTGTGTTATC 1551
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1552 ACGAGTGGATTTATCGACAAATGAGGGCAGACGGC 1587

RESULT 3

AAD28778 standard; cDNA; 1738 BP.

XX AAD28778;

XX 07-MAY-2002 (first entry)

XX Human 20P1F12-GTC1 cDNA.

XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTC1; cell growth; neoplasia;
XX cancer; vaccine; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
XX CDS 112..1591
XX /*tag= a
XX /product= "Human 20P1F12-GTC1 protein"

XX WO200204953-A2.

XX 17-JAN-2002.

XX 12-JUL-2001; 2001WO-US022168.

XX 12-JUL-2000; 2000US-00615285.

XX (AGEN-) AGENSYS INC.
XX Saferan D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
XX Challita-Eid PM;
XX WPI; 2002-154967/20.
XX P-PSDB; AAE18096.
XX Examining a biological sample for evidence of dysregulated cellular
XX growth, comprises comparing the status of prostate-specific, androgen-
XX regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
XX normal sample.
XX Example 3; Fig 1; 161pp; English.
XX The present invention relates to methods and compositions for the
XX diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX kidney cancer derived from or based on a normally prostate-specific,
XX androgen regulated, cell membrane associated secreted serine protease
XX termed 20P1F12/TMPRSS2. The invention further relates to a method of
XX examining a biological sample for evidence of dysregulated cellular
XX growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
XX designated 20P1F12-GTC1) in the sample to the status of 20P1F12/
XX in a corresponding normal sample. The invention also relates to 20P1F12/
XX TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
XX invention are used for examining a sample such as blood, serum, stool,
XX urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX The dysregulated cell growth is indicative of bladder cancer, lung
XX cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
XX expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
XX lung cancer, ovarian cancer or metastatic cancer. The present sequence is
XX human 20P1F12-GTC1 cDNA
XX Sequence 1738 BP; 433 A; 459 C; 473 G; 373 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 9, 14e-206 Length: 1738
Score: 2717.00 Matches: 492
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x AAD28778 (1-1738)

Qy 1 MetaLalaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 112 ATGGCTTTGAATCACTCAGGGTCCACCACCTATTGGACCTTACTATGAAACCATGGATAC 171
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 172 CAACCGGAACCCCTATCCGCACAGCCCACTGTGGTCCCACTGTCTACGAGTGCAT 231
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 232 CCGGCTCAGTACTACCCGCTCCCGCTCCCGACAGTACGCGGAGGCTCTGACGAGGCT 291
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 292 TCCAAACCCCGTCTGTGACGAGCCCAATCCCAATCCGACAGTGTGACCTCAAG 351
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
Db 352 ACTAAGAAGACACTGTGCATCACCTTGACCTCCGCGGAGCTTCTCTCTGGGAGCTCGCTG 411
Qy 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 412 GCCGCTGGCTACTCTGGAAGTTTCATGGGACAGAGTGTCCCAATCTGGGATAGAGTGC 471
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140


```
Db 472 GACTCTCAGTACCTGCATCAACCCCTCTAACTGGTGTGATGCGGTGCACACTGCCCC 531
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 532 GCGCGGGAGGACGAAATCGGTGTGTCGCTCTACGGACCAAACTTCATCCTTCAGGTG 591
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 592 TACTCATCTCAGAGGAAGTCTGGCACCTGTGTGCCAAGACGACTGGACGAGAACTAC 651
Qy 191 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 652 GCGCGGGCGGCTCGAGGACATGGCTATAGGAATAATTTTACTCTAGCAAGGAATA 711
Qy 201 ValAspSerSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 712 GTGGATGACAGGGATCCACAGCTTTATGAACCTGAACACAGTGGCGGCAATGTGAT 771
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 772 ATCTATAAAAACTGTACCACAGTGATGCTGTCTTCAAAGCAGTGGTTCCTTTACGC 831
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 832 TGTATAGCTCGGGGGTCACTTGACTCAGACCGCCAGACGAGATGTGGCGGCGAG 891
Qy 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 892 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGTCACTGACCTGACCTGCAGTCCAGAACGTCACGTG 951
Qy 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 952 TCGGAGGCTCCATCATCACCCCGAGTGGATCGTGCAGCGCCGACGCGCCCTGGGAAAA 1011
Qy 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1012 CCTCTTACATCCATGATGACGCACTTTCGGGGATTTTGAGCAATCTTTCATG 1071
Qy 321 PheTyrGlyAlaGlyTyrGlnValGlyLysValIleSerHisProAsnTyrAspSerLys 340
Db 1072 TTCTATGAGCGCGATACCAAGTAAAAAGTATTTCTCATCCAAATATGATCTCAAG 1131
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1132 ACCAAGAACATGACATTTGCGCTGATGAAGCTGCAGAACCTCTGACTTTCACAGACTA 1191
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1192 GTGAAACCAAGTGTGTCTGCCAACCCAGCGCATGCTGCAGCCAGAACAGCTCTGCTGG 1251
Qy 381 IleSerGlyTrpGlyAlaThrCluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1252 ATTTCCGGGTGGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTCTGACAGCTGCC 1311
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1312 AAGTGTCTTCTCATTGAGACACAGAGATGCACACAGCAGATATGTCTATGACAACTGATC 1371
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1372 ACACAGCAGCATGATCTGTGCGGCTCTCTGAGGGGAACGTCGATCTTCGACGGGTGAC 1431
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1432 AGTGAGGGCTCTGGTCACTTCGAGAACCAATATCTGTGTGCTCATAGGGGATACAGC 1491
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1492 TGGGGTCTCTGGTGTGCCAAAGCTTACAGACCAGAGTGTACGGGAATGTGATGTTATTC 1551
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
```

```
Db 1552 ACGGACTGGATTATCGACAAATGAGGGCAGACGGC 1587
RESULT 4
ADA10904
ID ADA10904 standard; cDNA; 2486 BP.
XX AC ADA10904;
XX DT 06-NOV-2003 (first entry)
XX DE Human cDNA differentially expressed in colon cancer #15.
XX KW ss; differential expression; colon cancer; cancer; human; gene.
XX OS Homo sapiens.
XX PN US2002160382-A1.
XX PD 31-OCT-2002.
XX PF 11-OCT-2001; 2001US-00981353.
XX PR 11-OCT-2000; 2000US-0239841P.
XX PA (LASE/) LASEK A W.
XX PA (JONE/) JONES D A.
XX PI Lasek AW, Jones DA;
XX DR WPI; 2003-265756/26.
XX DR P-PSDB; ADA10905.
XX PT New combination comprising cDNAs that are differentially expressed in
XX PT colon disorder, useful for diagnosing, treating, staging or monitoring
XX PS Claim 1; SEQ ID NO 22; 231bp; English.
XX CC The invention relates to a combination comprising cDNAs that are
XX CC differentially expressed in colon disorder. The methods and compositions
XX CC of the present invention are useful for diagnosing, treating, staging or
XX CC monitoring treatment for colon cancer. They are also useful in high
XX CC throughput methods for using cDNAs to detect differential expression of
XX CC nucleic acids in a sample, screening molecules or compounds to identify a
XX CC ligand which specifically binds a cDNA and using a protein to screen
XX CC molecules or compounds to identify at least one ligand which specifically
XX CC binds the protein. The present sequence represents a human cDNA
XX CC differentially expressed in colon cancer.
XX SQ Sequence 2486 BP; 564 A; 660 C; 719 G; 543 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,58e-205 Length: 2486
Score: 2712.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.82% Indels: 0
DB: Gaps: 0

US-09-615-285B-2 (1-492) x ADA10904 (1-2486)
Qy 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 149 ATGGCTTTGAACCTCAGGGTCACCCCTGCTATTTGGACCTTACTATGAAAACCATGGATAC 208
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValProThrValTyrGluValHis 40
Db 209 CAACCGGAAAAACCCCTATCCGCACAGCCACTGTGGTCCCCACTGTCTACGAGGTGCAT 268
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 269 CCGGCTCAGTACTACCGTCCCCCGTCCCGCCAGTACGCCCGGAGGTCTCTGACGCGGCT 328
```

QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 329 TCCAAACCCCGTCTGCTGACGACGACCAATATCCCATCCGGACAGTGTGCACCTCAAAG 388
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 389 ACTAAGAAAGCACTGTGCATCACCCTTGACCTGGGACCTTCTCTGGGAGCTGGCGTG 448
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 449 GCCCTGGCTACTCTGGAGTTCATGGGACGACAGTGTCTCCAACTCTGGAGTAGAGTGC 508
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 509 GACTCCTCAGTACCTGCATCAACCCCTTAACCTGGTGTGATGGCGTGTCCACTGCCCTCCC 568
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 569 GCGGGGAGGACGAGATCGGTGTGTCTGCTCTACGGACCAAACTTCATCTCTCAGGTG 628
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 629 TACTCATCTCAGAGAAAGTCTGCGACCTGTGTGCCAAGACGACTGGAAACGAGAACTAC 688
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 689 GCGCGGGCGGCTCGAGGACATCGGCTATAGAAATATTTTACTCTACCCAGGAATA 748
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 749 GTGGATGACAGCGGATCACCGATTTATGAACCTGAACCAAGTGGCGGCAATGTGCAT 808
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
Db 809 ATCTATAAAACCTGTACACAGATGCTGTCTTCTCAAAGCAGTGGTTCCTTACGC 868
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 869 TGTATAGCTCGGGGTCACTTGAACCTCAAGCCGCCAGAGAGATGTGGCGGCGAG 928
QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 929 AGCGCGTCCCGGGGCGCTGCGCTGCGAGCTGAGCTGCGAGAGTGTGGCGGCGAG 988
QY 281 CysGlyGlySerIleLeuThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 989 TCGGAGGCTCCATCATCACCCCGAGTGGACCGTGTACAGCCGCCACTCGGTGGAAAAA 1048
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1049 CCTCTTAACAATCCATGCAATTGACGGCAATTGGGGGATTTTGACACAATCTTTTCATG 1108
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1109 TTCTATGGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATTTATGACTCCAG 1168
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1169 ACCAAGACAATGACATTCGCTGATGAAGCTGCAGAGCCCTCTGACTTTCAACGACCTA 1228
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1229 GTGAACACCAAGTGTCTGCCCAACCCAGGATGATGTCTGCGCAGACAGCTCTGTCTGG 1288
QY 381 IleSerGlyTrpGlyValThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1289 ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTGTCTGAACGCTGCC 1348
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1349 AAGGTGCTTCTCATTTGACACACAGAGATGCAACAGCAGATATGTCTATGACAACTGATC 1408
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440

Db 1409 ACACGAGCCATGATCTGTGCGGCTTCTCGAGGGGAACGTCGATTCTTGCAGGGTGAC 1468
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1469 AOTGAGGGCTCTGCTCCTTCGAGAACCAATATCTGCTGCTGATAGGGGATACAAGC 1528
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1529 TGGGGTCTGCTGCTGCCAAAGCTTACAGACCAGAGTGTACGGGAATGTGTGTATTC 1588
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1589 ACGGACTGATTTATCGACAAATGAGGCGACAGCGC 1624
RESULT 5
ACH03998
ID ACH03998 standard; cDNA; 2486 BP.
XX ACH03998;
AC ACH03998;
XX 26-SEP-2003 (first entry)
DT Human cDNA differentially expressed in lung cancer #203.
DE Gene therapy; emphysema; ss; gene; chronic obstructive pulmonary disease;
KW respiratory disorder; lung cancer; asthma; human.
KW
XX Homo sapiens.
OS
XX US2003065157-A1.
XX
XX 03-APR-2003.
PD
XX 04-APR-2002; 2002US-00116802.
PF
XX 04-APR-2001; 2001US-0281593P.
PR
XX (LASE/) LASEK A W.
PA
XX Lasek AW;
PI
XX WPI; 2003-540803/51.
DR
XX New combination comprising cDNAs that are differentially expressed in
PT respiratory disorders, useful for diagnosing or treating respiratory
PT disorders e.g., lung cancer, chronic obstructive pulmonary disease,
PT emphysema or asthma.
XX
XX Claim 1; Page; 39pp; English.
XX
XX The invention relates to a combination comprising cDNAs or their
CC complements that are differentially expressed in respiratory disorder.
CC The combination is useful for preparing a composition for diagnosing or
CC treating respiratory disorders e.g. lung cancer, chronic obstructive
CC pulmonary disease, emphysema or asthma. The present sequence represents
CC human cDNA differentially expressed during lung cancer
XX
SQ Sequence 2486 BP; 564 A; 660 C; 719 G; 543 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 3 58e-205 Length: 2486
Score: 2712.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.82% Indels: 0
DB: Gaps: 0
US-09-615-285B-2 (1-492) x ACH03998 (1-2486)
QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 149 ATGGCTTTGAACCTCAGGTCACCACTGCTATTTGACCTTACTATGAAACCAACCATGATAC 208

QY 21 GlnProGluAAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 209 CACCGGAAACCCCTATCCCGCAGCCCACTGTGGTCCCACTGTCTAGAGGTGCAT 268
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 269 CCGGCTCAGTACTACCGGTCCCGCTGCCAGTAGTACGCCCGAGGGTCTTGACGAGGCT 328
 QY 61 SerAAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 329 TCCAAACCCGCTGTCTGACGAGCCCAATCCCACTCCGGACAGTGTGACCTCAAG 388
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 389 ACTAAGAAGCACTGTGCATCACCTTGACCTGGGACCTTCTCTGCTGGAGCTGGCGTG 448
 QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAAsnSerGlyIleGluCys 120
 Db 449 GCGGCTGGCTACTCTGGAAGTTCATGGCGACAGCAAGTCTCCAACTCTGGGATAGAGTGC 508
 QY 121 AspSerSerGlyThrCysIleAAsnProSerAAsnTyrCysAspGlyValSerHisCysPro 140
 Db 509 GACTCCTCAGGTACTGTGCATCAACCCCTCTAACTGGTGTGATGGCGTGTCACTGCCCC 568
 QY 141 GlyGlyGluAspGluAAsnArgCysValArgLeuTyrGlyProAAsnPheIleLeuGlnVal 160
 Db 569 GCGGGGAGGACAGAAATCGGTGTCTTGCCTCTACGGACCAAACTTCTCATCTTCAAGTG 628
 QY 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
 Db 629 TACTCATCTCAGAGAAGTCTCGGCACCTGTGTGCCAAGACGACTGGAACGAACTAC 688
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAAsnAAsnPheTyrSerSerGlnGlyIle 200
 Db 689 GCGCGGGCGGCTCGAGGACATGGCTATTAAGAATATTTTACTCTAGCAAGGAATA 748
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAAsnThrSerAlaGlyAAsnValAsp 220
 Db 749 GTGGATGACAGCGGATCCACACAGCTTTATGAACCTGAACACAGTGGCGGCAATGTGCAT 808
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 809 ATCTATAAAACTGTATCACAGTGTGCTGCTTCTTCAAAAGCAGTGGTTCTTTACGC 868
 QY 241 CysIleAlaCysGlyValAAsnLeuAAsnSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 869 TGTATAGCTTGGGGGTCAACTTGAACCTCAAGCCGCCAGACAGCAGGATTGTGGCGGGAG 928
 QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAAsnValHisVal 280
 Db 929 AGCGGCTCCCGGGGCTTGGCCCTGGCAGTGCAGCTGCAGCTCCAGAACGTCCACGTG 988
 QY 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
 Db 999 TCGGAGGCTCCATCATCACCCCGGAGTGGACCGGTGACAGCGGCCCATCTGGTGGAAAA 1048
 QY 301 ProLeuAAsnAAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 1049 CCTCTTAACAATCCATGGCATTTGGCGGATTTGGCGGATTTTGAGCAATCTTTTCATG 1108
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAAsnTyrAspSerLys 340
 Db 1109 TTCTATGAGCGCGGATACCAAGTAGAAGAGTATTTCTCATCCAAATATATGACTCCAAG 1168
 QY 341 ThrLysAAsnAAsnIleAlaLeuMetLysLeuGlnLysProLeuThrPheAAsnAspLeu 360
 Db 1169 ACCAAGAACAATGACATTTGCGTGTGAAGCTGCAGAGCCCTCTGACTTTTCAAGCACTA 1228
 QY 361 ValLysProValCysLeuProAAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 Db 1229 GTGAACCAAGTGTGTCTGCCCAACCCAGGCAATGATGCTGCAGCCAGAACACTCTGTGG 1288

QY 381 ILeSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAAsnAlaAla 400
 Db 1289 ATTTCCGGGTGGGGGCCACCCAGGAGAAAGGAAGACCTCAGAGTGTCTGAACGCTGCC 1348
 QY 401 LysValLeuLeuLeuLeuGluThrGlnArgCysAAsnSerArgTyrValTyrAspLeuLeu 420
 Db 1349 AAGGTGCTTCTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGACACCTGATC 1408
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAAsnValAAspSerCysGlnGlyAsp 440
 Db 1409 ACACCGACCATGATCTGTGCGGCTTCTGCAGGGAAGCTCGATTCTTGCACGGGTGAC 1468
 QY 441 SerGlyGlyProLeuValThrSerLysAAsnAAsnIleTyrTyrLeuLeuGlyAspThrSer 460
 Db 1469 AGTGGAGGGCTCTGTGCTCACTTCGAAGAACAATATCTTGGTGGCTGATAGGGGATCAAGC 1528
 QY 461 TrrGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAAsnValMetValPhe 480
 Db 1529 TGGGGTGTGGCTGTGCCAAAGCTTACAGCCAGGAGTGTACGGGAATGTGATGTATTC 1588
 QY 481 ThrAspTrrPheTyrArgGlnMetArgAlaAspGly 492
 Db 1589 ACGGACTGGATTTATCGAACAATGAGGGCAGACGGC 1624

RESULT 6

ADB75590
 ID ADB75590 standard; cDNA; 3226 BP.

AC ADB75590;

DT 04-DEC-2003 (first entry)

XX Prostate cancer marker cDNA.

DE Prostate; cancer; cytostatic; gene therapy; marker; ss.

KW Homo sapiens.

XX WO2003009814-A2.

PN 06-FEB-2003.

PD 25-JUL-2002; 2002WO-US023913.

PF 25-JUL-2001; 2001US-0307982P.

PR 22-AUG-2001; 2001US-0314356P.

PR 25-SEP-2001; 2001US-0325020P.

PR 12-DEC-2001; 2001US-0341746P.

PR 05-MAR-2002; 2002US-0362158P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B;

PI Hoerish S, Kamatkar S, Woney AM, Glatt K, Zhao X, Anderson D;

XX WPI; 2003-248033/24.

XX New nucleic acid molecule, useful for diagnosing or treating prostate cancer.

XX Disclosure; SEQ ID NO 414; 99pp; English.

CC The invention relates to newly discovered cancer markers associated with the cancerous state of prostate cells. Also disclosed is a method of assessing whether a patient is afflicted with prostate cancer. The method of the invention involves assessing whether a patient is afflicted with prostate cancer by comparing the level of expression of a marker in a patient sample and the normal level of expression of the marker in a control non-prostate cancer sample, where a significant increase in the level of expression of the marker in the patient sample and the normal level indicates that the patient is afflicted with prostate cancer. Nucleic acids of the invention are useful for diagnosing or treating prostate cancer, and may be useful in gene therapy. Sequences given in

CC ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences.

XX SQ Sequence 3226 BP; 782 A; 808 C; 857 G; 779 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5, 96e-205 Length: 3226
Score: 2711.00 Matches: 490
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x ADB75590 (1-3226)

QY	1	MetAlaLeuAenSerGlySerProAlaIleGlyProTyrTyrGluAenHisGlyTyr	20
DB	128	ATGGCTTTGAACCTCAGGCTACCCAGCTATTGGACTTACTATGAAACCCATGGATAC	187
QY	21	GlnProGluAenProTyrProAlaGlnProThrValProThrValTyrGluValHis	40
DB	188	CAACCGGAACCCCTATCCCGCAGACCCACTGTGTCCCACTGTCTACGAGGTGCAT	247
QY	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
DB	248	CCGGCTCAGTACTACCCCTCCCGTCCCGCAGTACCGCCGAGGGTCCCTGACGCGAGCT	307
QY	61	SerAenProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
DB	308	TCCAAACCCCGTCTGTCACACGACGCCAATCCCTCCCGGACAGTGTGCACCTCAAG	367
QY	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
DB	368	ACTAAGAAAGCACTGTGCATCACCTTGACCTGGGGACCTTCTCTCGTGGGAGCTGGCTG	427
QY	101	AlaAlaGlyLeuLeuTyrPhePheGlySerLysCysSerAenSerGlyIleGluCys	120
DB	428	CCCGTGGCCCTACTCTGGAAGTTTATGGGAGAGAGTCTCCAACTCTGGATAGAGTGC	487
QY	121	AspSerSerGlyThrCysIleAenProSerAenTrpCysAspGlyValSerHisCysPro	140
DB	488	GACTCTCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCTGCACACTGCCCC	547
QY	141	GlyGlyGluAenAspGluAenArgCysValArgLeuTyrGlyProAenPheIleLeuGlnVal	160
DB	548	GGCGGGAGGACGAGAAATCGGTGTGTTCGCCCTCTACGGACCAAACTTCATCTCTCAGATG	607
QY	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAenGluAenTyr	180
DB	608	TACTCATCTCAGAGGAATCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGACTAC	667
QY	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAenAenPheTyrSerSerGlnGlyIle	200
DB	668	GGCGGGCGGCTGCAAGGACATGGGCTTATAGGAATAATTTTACTCTAGCCAAAGGAATA	727
QY	201	ValAspAspSerGlySerThrSerPheMetLysLeuAenThrSerAlaGlyAenValAsp	220
DB	728	GTGGATGACGCGGATCCACAGCTTTATGAATCGAACACAAGTGGCGCAATGTGCAT	787
QY	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
DB	788	ATCTATAAAAACTGTACACAGTGTGCTGCTTCTTCAAAAGCAGTGGTTTCTTTACGC	847
QY	241	CysIleAlaCysGlyValAenLeuAenSerSerArgGlnSerArgIleValGlyGlyGlu	260
DB	848	TGTATAGCTCGGGGGTCAACTTGAACCTCAGCGCCGACAGAGATTGGGGCGGTGAG	907
QY	261	SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAenValHisVal	280
DB	908	AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGGTTCAGCTGCACGTCCAGAACCTCCACGTG	967

QY	281	CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys	300
DB	968	TGCGGAGGCTCCATCATCACCCCCGAGTGGATCGTACAGCGGCCCACTGGTGGAAAA	1027
QY	301	ProLeuAenAenProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
DB	1028	CCTCTTAACAATCCATGGCATTTGGACGATTTGGGGGATTTTGAGCAATCTTTTCATG	1087
QY	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAenTyrAspSerLys	340
DB	1088	TTCTATGAGCCGGATACCAAGTAGAAAAAGTATTCTCATCCAAATATTGACTCCAAG	1147
QY	341	ThrLysAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu	360
DB	1148	ACCAAGAACAATGACATTCGCTGATGAAGCTGCAAGAGCTCTGACTTTCAACGACCTA	1207
QY	361	ValLysProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTrp	380
DB	1208	GTGAACCCAGTGTGTCTGCCCAACCCAGGCATGCTGCAGCCAGAACAGCTCTGCTGG	1267
QY	381	IleSerGlyTyrGlyAlaThrGluGlyLysThrSerGluValLeuAenAlaAla	400
DB	1268	ATTTCGGGTGGGGGCCACCCAGAGAGAAAGGAGACCTCAGAAAGTGTGAACCTGCC	1327
QY	401	LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrValTyrAspAenLeuIle	420
DB	1328	AAAGTGTCTCTCTTGGACACACAGATGCCACAGACATATGTCTATGACACCTGATC	1387
QY	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAenValAspSerCysGlnGlyAsp	440
DB	1388	ACACCCAGCCATGATCTGTCCCGCTTCTGCAGGGGAAACGTGATTTCTTGCAGGGTGAC	1447
QY	441	SerGlyGlyProLeuValThrSerLysAenAenIleTrpTrpLeuIleGlyAspThrSer	460
DB	1448	AGTGGAGGCGCTCTGGTCACTTCGAGAAACAATATCTGGTGGCTGATAGGGATACAGC	1507
QY	461	TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAenValMetValPhe	480
DB	1508	TGGGGTCTCGCTGTGCCAAAGCTTACAGACCAGAGTGTACGGGAATGTGATGTATTTC	1567
QY	481	ThrAspTrpIleTyrArgGlnMetArgAlaAspGly	492
DB	1568	ACGGACTGGATTTATCGACAAATAGGGCGACAGCGC	1603

RESULT 7

AAH93944
ID AAH93944 standard; cDNA; 1476 BP.
XX AC AAH93944;
XX DT 04-OCT-2001 (first entry)
XX DE P1000C open reading frame cDNA sequence without stop codon.
XX KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
XX KW cytostatic; gene therapy; metastasis; ss.
XX OS Homo sapiens.
XX PN WC200151633-A2.
XX PD 19-JUL-2001.
XX PF 16-JAN-2001; 2001WO-US0001574.
XX PR 14-JAN-2000; 2000US-00483672.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX

DR WPI; 2001-425873/45.
XX New polynucleotide encoding a prostate-specific protein, for diagnosing,
PT monitoring and treating prostate cancer in a patient and for use in
PT vaccines.
XX
XX
PS Claim 1; Page 538-539; 543pp; English.

XX The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (I), fusion proteins comprising (II), and isolated T cells
CC prepared using (I) or (II) are used to treat cancer in a patient. (I) and
CC the antibodies are also used in the detection of cancer in a patient. The
CC cancer that is diagnosed or treated is particularly prostate cancer. (I)
CC and (II) can be used in vaccines. The antibodies or (I) can be used for
CC monitoring the progression of cancer in a patient. (I) and (II) can also
CC be used to improve diagnostic and therapeutic methods for prostate
CC cancer. They can indicate the level of metastasis as well as the prostate
CC volume. AAH93357 to AAH93944 and AAH0115 to AAH01318 represent
CC polynucleotide and amino acid sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 1476 BP; 361 A; 405 C; 402 G; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3, 21e-205 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
Gaps: 0

US-09-615-285B-2 (1-492) x AAH93944 (1-1476)

QY	1	MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
DB	1	ATGGCTTTGAACCTCAGGCTACCCAGCATTGGACCTTACTATGAACCATGGATAC	60
QY	21	GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis	40
DB	61	CAACCGGAAACCCCTATCCCGCAGCCACTGTGGTCCCACTGTCTCAGAGTGCAT	120
QY	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
DB	121	CGGGCTCAGTACTACCCGTCCTCCCGCCAGTAGCCGCCGAGGGTCTCTGACGAGGCT	180
QY	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
DB	181	TCCAAACCCCGTCTGTGCAGCAGCCCAATCCCAATCCCGACAGTGTGCACCTCAAAG	240
QY	81	ThrLysLysAlaLeuCysAlaThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
DB	241	ACTAAGAAAGCACTGTGCATCACTTGACCTGGGGACCTTCTCTGGGGAGCTGGCTG	300
QY	101	AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyLeuGluCys	120
DB	301	GCCTGTGGCTACTCTGGAAGTTCATGGGACAGAAAGTGTCTCAACTCTGGGATAGAGTGC	360
QY	121	AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro	140
DB	361	GACTCCTCAGGTACTGTGCATCAACCCCTCTAACTGTGTGTGGGTGTGCACCTGCCCC	420
QY	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal	160
DB	421	GCCTGGGAGGACGAGATCGGTGTGTCTGCTCTACGGATCAAACTTCTCCTTCAGGTG	480
QY	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr	180
DB	481	TACTCATCTCAGAGGAAGTCTGGGACCCCTGTGTGCCAAGACGACTGGAAACGAGACTAC	540
QY	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200

DB	541	GGCGGGCGGCGCTGCAGGACATGGGCTATAGATAAATTTTACTCTAGCCAGGAATA	600
QY	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
DB	601	GTGGATGACAGCGGATCCACCGAGCTTTATGAAATGAACTGAACACCAAGTCCCGCAATGTCAT	660
QY	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
DB	661	ATCTATATAAAATCTGTACCAAGTGTCTTCTTCAAAAGCAGTGGTCTTTTTCACGC	720
QY	241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu	260
DB	721	TGTATAGCTCGGGGTCAACTTGAATCAAGCCGCCAGAGCAGGATTGTGGCGGCGAG	780
QY	261	SerAlaLeuProGlyAlaIleProProTrpGlnValSerLeuHisValGlnAsnValHisVal	280
DB	781	AGCGCGCTCCCGGGGCGCTGGCCCTGCAGGTGAGCTGCAGTCCAGAACGTCACAGTG	840
QY	281	CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys	300
DB	841	TGGGAGGCTCCATCATCAACCCGAGTGTGATCGTACAGCCGCCCTGCTGGTGAATAA	900
QY	301	ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
DB	901	CCTCTTAACAATCCATGGCATTCGACGGCATTTGCGGGGATTTTGAGACAAATCTTTCATG	960
QY	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys	340
DB	961	TTCTATGGAGCCGATACCAAGTAGAAAGTAGTATTTCTATCCAAATATATGACTCCAG	1020
QY	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
DB	1021	ACCAAGAACAAATGACATTGCGCTGTGTAAGTGTGAGAGCCTCTGACTTTCAACGACCTA	1080
QY	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGlnLeuLysTyr	380
DB	1081	GTGAACACCAAGTGTGTCTGCCCAACCCAGGACATGATCTGCAGCCAGACAGCTCTCTGG	1140
QY	381	IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla	400
DB	1141	ATTTCCGGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAGTGTGTGAACGCTGCC	1200
QY	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle	420
DB	1201	AAAGTGTCTTCTCATTCAGACACAGATGCACAGCAGATATGCTATGACAACTGATC	1260
QY	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp	440
DB	1261	ACACCGCCCATGATCTGTCCGGCTTCTTCAGGGGAACTGCTGATTTCTTGGCAGGCTGAC	1320
QY	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer	460
DB	1321	AGTGGAGGCGCTCTGGTCACTTCGAAGAACAAATATCTGGTGGCTGTAGGGGATACAAGC	1380
QY	461	TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe	480
DB	1381	TGGGGTCTGGCTGTGCCAAAGCTTACAGACCCAGGAGTGTACGGGAATGTGATGTTATTC	1440
QY	481	ThrAspTrpIleTyrArgGlnMetArgAlaAspGly	492
DB	1441	ACGGACTGGATTTATCGACAAATGAGGGCAGACCGC	1476

RESULT 8
AAS64180
ID AAS64180 standard; cDNA; 1476 BP.
XX
XX AAS64180;
XX
XX 29-JAN-2002 (first entry)
XX
XX Human prostate cDNA sequence #599.
XX
XX Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.
KW

XX Homo sapiens.
XX WO200173032-A2.
XX 04-OCT-2001.
XX 27-MAR-2001; 2001WO-US009919.
XX 27-MAR-2000; 2000US-00536857.
XX 09-MAY-2000; 2000US-00568100.
XX 12-MAY-2000; 2000US-00570737.
XX 13-JUN-2000; 2000US-00593793.
XX 27-JUN-2000; 2000US-00605783.
XX 09-AUG-2000; 2000US-00636215.
XX 29-AUG-2000; 2000US-00651236.
XX 06-SEP-2000; 2000US-00657279.
XX 02-OCT-2000; 2000US-00679426.
XX 10-OCT-2000; 2000US-00685166.
XX 09-NOV-2000; 2000US-00709729.
XX (CORI-) CORIXA CORP.
XX Xu J, Dillon DC, Mitcham JL, Harlocker SH, Jiang Y, Kalos MD;
XX Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
XX Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX WPI: 2001-539232/73.
XX F-PSDB; AAU69960.
XX New human prostate-specific polypeptides and polynucleotides useful for
XX the diagnosis and treatment of cancer, especially prostate cancer.
XX Claim 1; Page 572; 579pp; English.
XX The invention relates to isolated prostate-specific polynucleotides,
XX polypeptides, fusion proteins of the polypeptides, antibodies raised
XX against the polypeptides (or antigenic epitopes derived from them) and
XX antigen-presenting cells expressing the polypeptides. The antibodies are
XX useful for detecting the presence of cancer, especially prostate cancer.
XX The polypeptides, polynucleotides and the antigen-presenting cells are
XX useful for stimulating and/or expanding T cells specific for a tumour
XX protein, and for inhibiting the development of cancer especially prostate
XX cancer. Compositions comprising the polynucleotide and/or polypeptide are
XX useful for stimulating an immune response, and for treating cancer. The
XX oligonucleotide is useful for detecting cancer. The present sequence is a
XX prostate specific polynucleotide of the invention
XX SQ Sequence 1476 BP; 361 A; 405 C; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,218-205 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 4 Gaps: 0

US-09-615-285B-2 (1-492) x AAS64180 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 1 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGAAACCATGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 61 CAACCGGAAACCCCTATCCCGCAGACCCACTGTGGTCCCACTGCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 121 CCGGCTCAGTACTACCTCCCTCCCGTGCACGACGAGGAGGCTCTGACGCGAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80

DB 181 TCCAAACCCCGTCTGCAACGACGCCAAATCCCACTCCGGGACAGTGTGACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 241 ACTAAGAAAGCACTGTGTCATCACTTACCCTGGGGACCTTCTCTGCTGGGAGTGCCTG 300
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 301 GCCGCTGGCCTACTCTGGAAGTTTATGGGACGAGTGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AppSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
DB 361 GACTCTCTCAGGTACTGTCATCAACCCCTCTAACTGGTGTGATGGCGTGTCCACACTGCC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 421 GGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspPheTrpAsnGluAsnTyr 180
DB 481 TACTCATCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 541 GGGGGGGGGCTGCAGGGGACATGGGCTATAAGAAATAATTTTACTCTAGCAAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 601 GTGGATGACACGGGATCCACGAGCTTTATGAACTGAACACCAAGTCCCGCAATGTGAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAANAACGTGTACCACTGATGCTGTCTTCAAAAGCAGTGGTCTTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
DB 721 TGTATAGCTCGGGGTCAACTTGAACCTCAAGCGCCGAGCAGGAGTGTGGCGCGGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 AGCGGCTCCCGGGGGCTGCGCTGCGAGTCACTGAGCTGACGAGTCCAGAACGTCACGTG 840
QY 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
DB 841 TGGGAGGCTCCATCATCACCCTGAGTGTGATGTCGACAGCCGCCCACTGCTGGGAAAA 900
QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CCTCTTAACAATCCATGGCATTCGAGCGCATTTTGGGGGATTTTGAGACAATCTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
DB 961 TTCTATGGAGCGGATACCAAGTAGAAAAAGTAGTATTTCTCATCAAATATATGACTCCAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
DB 1021 ACCAAGAACAATGACATTCGCTGATGAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGG 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGlnGlnLeuCysTrp 380
DB 1081 GTGAACCCAGTGTGTCTGCCCAACCCAGGATGATGTCGAGCCAGCAACAGACTCTCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
DB 1141 ATTTCCGGGTGGGGGGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
DB 1201 AAGTGTCTTCTCATTTAGACACAGAGATGCAACAGCAGATATGCTATGACACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440

Qy 221 IletyrlslyslsleuTyHsSerAspAlaCysSerSerlyeAlaValValSerLeuAig 240
Db |||||
241 CysileAlaCysGlyValAenLeuAenSerSerArgInSerArgIleValGlyGlyGlu 260
Db |||||
721 TGTATAGCTTGGCGGGTCAACTTGAATCAAGCCGCGAGAGGAGTGTGGCGCGCGAG 780
Qy 261 SerAlaLeuProGlyAlaTatPProTtPInValSerLeuHisValGlnAenValHisVal 280
Db |||||
781 AGCGCGCTCCCGGGGGCTGGCTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 840
Qy 281 CysGlySerIleIleThrProGluTtPileValThrAlaAlaHisCysValGlyLys 300
Db |||||
841 TGGGAGGCTCCATCATCACCCCGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 900
Qy 301 ProLeuAenAenProTtHsTtPThrAlaPheAlaGlyIleLeuArgInSerPheMet 320
Db |||||
901 CCTTTAAACATCATGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 960
Qy 321 PheTyrlGlyAlaGlyTyrlGlnValGlyLysValIleSerHisProAenTyrlAspSerLys 340
Db |||||
961 TTCTATGAGCGGATACCAAGTAGAAAGTAGATTTCTCATCAATATGACTCCAAG 1020
Qy 341 ThrlyAenAenAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAenAspLeu 360
Db |||||
1021 ACCAAGAACATGATGCTGGCTGATGAAGTGGAGTGGAGTGGAGTGGAGTGGAG 1080
Qy 361 VallylsProValCysLeuProAenProGlyMetMetLeuGlnProGluGlnLeuCysTtp 380
Db |||||
1081 GTGAACACAGTGTCTGCCACCCAGGATGATGCTGGAGGAGGAGGAGGAGGAGG 1140
Qy 381 IleSerGlyTtpGlyAlaThrGluGluGlyGlyLysThrSerGluValLeuAenAlaAla 400
Db |||||
1141 ATTTCCGGTGGGGGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAenSerArgTyrlValTyrlAspAenLeuIle 420
Db |||||
1201 AAGGTGCTTCTATGAGACACAGATGCAACAGCAGATATGCTATGACACCTGATC 1260
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlnValAenValAspSerCysGlnGlyAsp 440
Db |||||
1261 ACACGACCATGATCTGTGGCGGCTTCTGTGAGGAGGAGGAGGAGGAGGAGGAG 1320
Qy 441 SerGlyProLeuValThrSerLysAenAenIleTtPThrLeuIleGlyAspThrSer 460
Db |||||
1321 AGTGGGGGGCTTGTGGTCACTTCGAAGAACAAATCTGGTGGCTGATAGGGGATACAAGC 1380
Qy 461 TtpGlySerGlyCysAlaLysAlaTyrlArgProGlyValTyrlGlyAenValMetValPhe 480
Db |||||
1381 TGGGTTCTGGCTGTGCAAGCTTACAGACGAGGAGTGTACGGAATGTGATGTATTTC 1440
Qy 481 ThrAspTtPleTyrlArgGlnMetArgAlaAspGly 492
Db |||||
1441 ACGGACTGGATTTATCGAACAAATGAGGGCAGACGGC 1476

RESULT 10

ABL95551 standard; cDNA; 1476 BP.

AC ABL95551;

19-JUL-2002 (first entry)

Human P100C cDNA sequence SEQ ID NO 931.

Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;
Gene therapy; gene; ss.

Homo sapiens.

US2002022248-A1.

XX

PD 21-FEB-2002.
XX
PF 12-JAN-2001; 2001US-00759143.
XX
PR 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 10-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.
PR 09-APR-1999; 99US-00288946.
PR 13-JUL-1999; 99US-00352816.
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 10-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.

XX (XUJJ/) XU J.

PA (DILL/) DILLON D C.

PA (MITC/) MITCHAM J L.

PA (HARL/) HARLOCKER S L.

PA (JIAN/) JIANG Y.

PA (KALO/) KALOS M D.

PA (FANG/) FANGER G R.

PA (RETT/) RETTER M W.

PA (STOL/) STOLK J A.

PA (DAYC/) DAY C H.

PA (VEDV/) VEDVICK T S.

PA (CART/) CARTER D.

PA (LISX/) LI S X.

PA (WANG/) WANG A.

PA (SKEI/) SKEIKY Y A W.

PA (HEPL/) HEPLER W T.

PA (HEND/) HENDERSON R A.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;

PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;

PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;

XX WPI; 2002-255649/30.

XX New prostate-specific polynucleotides for diagnosing and treating

XX diseases, in particular prostate cancer, and as markers for the

XX progression of cancer.

XX Claim 1; SEQ ID NO 931; 87pp; English.

XX The present invention provides prostate-specific coding sequences and

XX their encoded proteins. These can be used in the diagnosis and treatment

XX of cancers, particularly prostate cancer. The present sequence is a cDNA

XX described in the invention

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Alignment Scores:

Pred. No.: 3.21e-205 Length: 1476

Score: 2709.00 Matches: 491

Percent Similarity: 99.80% Conservative: 0

Best Local Similarity: 99.80% Mismatches: 1

Query Match: 99.71% Indels: 0

DB: Gaps: 0

US-09-615-285B-2 (1-492) x ABL95551 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 1 ATGGCTTTTGAACATCAGGGTCACACCAGCTATTGGACCTTACTATGAAAAACCATGATAC 60
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 Db 61 CAACCGGAACACCCCTATCCGACACAGCCCATGTGGTCCCACTTCTACAGATGTCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CCGGCTCAGTACTACCCGTCCTCCCGTCGCCAGTAGTCGCCCGGCTCTGACGAGGCT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCCAAACCCCGTGTCTGSCACGACGCCCAATCCCATCCGACAGTGTGCACCTCAAAG 240
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 241 ACTAGAAAGCACTGTCATCACTTGACCTGGGACCTTCTCTGTGGAGCTCGGCTG 300
 QY 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 301 GCCGCTGGGCTACTCTCGAAGTTCATGGGCAGCAAGTCTCCAACTCTGGGATAGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
 Db 361 GACTCTCAGTACCTGCACTCAACCCCTCTAACTGTGTGATGGGTGTCACTGCTGCC 420
 QY 141 GlyGlyLeuAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 421 GCGGGGAGGAGCAGAAATCGGTGTGTGCTCTACGGATCAAACTTCATCTCTCAGGTG 480
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 Db 481 TACTCATCTCAGAGGAAGTCTGGCACCTGTGTGCCAAGACGACTGGAACGAGAACATAC 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 541 GGGCGGGCGCTGTCAGGACATGGCTATAGATAATATTTTACTTAGCCACAGGATA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACCAAGCTTTATGAAACTCAACACAAAGTCCCGCAATTCGAT 660
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 661 ATCTATAAAACCTGTACCACAGTGTATGCTGTCTTCAAAAGCAGTGTGTCTTTACGC 720
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgInSerArgIleValGlyGlyGlu 260
 Db 721 TGTATAGCTGCGGGGTCAACTTGAACTCAAGCGCCAGCAGCAGGATTGTGGCGCGCAG 780
 QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 781 AGCGGCTCCCGGGGCGCTGGCCCTGGCAGCTGCAGCTGCAGAACGCTCCACGTG 840
 QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
 Db 841 TGGGAGGCTCCATCATCACCCCGAGTGTGATCGTGACAGCGCCCACTCGGTGAAAAA 900
 QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 901 CCTCTTTAAACATCATGAGCATTTGGACGGCATTTGGCGGGATTGTGAGACAATCTTTTCATG 960
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 961 TTCTATGGAGCCCGATACCAAGTAGAAAAGTGAATTTCTCATCCCAATATGACTCCAAAG 1020
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1021 ACCAAGAACAAATGACATTCGCTGTATGAAGCTGCAGAGCCCTGCACTTTCAACGACCTTA 1080

QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 Db 1081 GTGAACACAGTGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGAACAGCTCTGCTGG 1140
 QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 Db 1141 ATTTCCCGGGTGGGGGCCACCGAGGAGAAAGGAACCTCAGAAAGTCTGAACCTGCC 1200
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1201 AAGTGTCTTCTCATTTAGACACACAGAGATGCAACAGCAGATATGTCTATGACCAACTGATC 1260
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1361 ACACGACCCATGATCTGTCCGGCTTCTCGAGGGAACTCGGATCTTTCGCCAGGGTGAC 1320
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpIleLeuIleGlyAspThrSer 460
 Db 1321 AGTGAGGGCTCTGTGTCACCTTCGAAGAACAATATCTGTGTGGCTGATAGGGGATACAAGC 1380
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1381 TGGGGTTCTGGCTGTGCCAAAGCTTACAGACCAGGAGTGACGGGAATGTGATGGTATTC 1440
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 Db 1441 ACGGACTGGATTATCGACAAATGAGGGCAGACGGC 1476
 RESULT 11
 ACC95715
 ID ACC95715 standard; cDNA; 1476 BP.
 XX
 AC ACC95715;
 XX
 DT 28-AUG-2003 (first entry)
 XX
 DE Prostate tumour specific cDNA sequence SEQ ID 931.
 KW
 KW Cytostatic; gene therapy; prostate-specific protein; PSP; human;
 immune response; prostate cancer; ss.
 XX
 OS Homo sapiens.
 XX
 FN WO200289747-A2.
 XX
 PD 14-NOV-2002.
 XX
 PF 09-MAY-2002; 2002WO-US014753.
 XX
 PR 09-MAY-2001; 2001US-00852911.
 PR 29-JUN-2001; 2001US-00895814.
 PR 10-DEC-2001; 2001US-00012896.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Henderson RA;
 PI Kalos MD, Fanger GR, Retter MW, Stoik JA, Day CH, Vedvick TS;
 PI Carter D, Li SX, Wang A, Skeiky YAW, Hepler WT, Hural J;
 PI McNeill PD, Houghton RL, Vinals Y De BassolaC, Foy TW, Watanabe Y;
 PI Deng T;
 XX
 DR WPI; 2003-167130/16.
 XX
 PT New prostate-specific proteins and genes, useful in gene therapy,
 PT particularly for stimulating an immune response in a patient, or treating
 PT prostate cancer in a patient, as well as for diagnosing prostate cancer
 PT in a patient.
 XX
 PS Example 2; Page 642; 691pp; English.
 XX
 CC The present invention relates to novel prostate-specific proteins (PSP)
 CC and their coding sequences. The PSPs and their coding sequences are
 CC useful for stimulating an immune response in a patient, or for treating

CC prostate cancer in a patient and for determining, detecting or diagnosing
CC the presence of a cancer in a patient. The present sequence was used to
CC illustrate the invention

XX Sequence 1476 BP; 361 A; 405 C; 402 G; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,21e-205 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
Dbs: 7 Gaps: 0

US-09-615-285B-2 (1-492) x ACC95715 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAATCAGGCTCACCACAGCTATTGGACCTTACTATGAAACCATGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCCGCAGCCCACTGTGTCCCATCTGTCTACGAGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCGCCCGTGCACCAGTACGCCCGAGGGTCTTGACGAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCCGCTGTGTGACGAGCCCAATCCCATCCCGGAGAGTGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAGAAAGCACTGTGCATCACCTTGACCTGGGACCTTCCTCGTGGAGCTGCCTG 300
QY 101 AlaAlaGlyLeuLeuTyrPhePheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCCGCTGGCTACTCTGGAAGTTATGGGCGACAGTGTCTCCAACTCTGGGATAGATGC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 361 GACTCCTCAGGTAAGTCTGATCAACCCCTCTACTGGTGTGTGATGGTGTGCACCTGC 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 421 GCGGGGAGGACGAGATCGGTGTGTGCTCTACCGATCAAACTTCATCTTCAGGTG 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGAAAGTCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGAATAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GGGCGGGCGCTCTCAGGACATGGCTATAGAATAATTTTACTCTAGCCAAAGGAATA 600
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAACATGAACACAAAGTGGCGGAATGTCAT 660
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAACTGTACACAGTGTGCTCTTCTTCAAAAGCAGTGTGTTCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 721 TGTATAGCTGGGGTCAACTGAACCTCAAGCCGACAGCAGGATTTGTGGCGGCGAG 780
QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGGCTCTCCGGGGCCCTGGCCCTGGCAGTCAAGCTGCAGTCCAGAACGTCACCTG 840
QY 281 CysGlyGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300

Db 841 TCGGAGGCTCCATCATCATCCCCCGAGTGGATCGTGACAGCGCCCACTCGGTGAAAAA 900
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAAATCCATGCGCATTTGGACGCAATTTGGGGGATTTGAGACAAATCTTTCATG 960
QY 321 PheTyrGlyValaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCATGAGCGCGATACCAAGTAGAAAAAGTGAATTTCTCATCAAAATTTATGACTCAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGACATTTGGCTGATGAAGCTGCAAGAGCCCTCTGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1081 GTGAAACCAAGTGTGTCTGCCAACCCAGCATGATGCTGCAGCCAGACAGCTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCGGGTGGGGGCCACCGAGGAGAAAGGAGAACCTCAGAAGTGTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGGTCTTCTCATTTGAGACAGAGATGCAACAGCAGATATGTCTATGACACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACACGACCATGATCTGTGCGGCTTCTGTGAGGGGAACGTCGATTTCTGCCAGGGTAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpIleGlyAspThrSer 460
Db 1321 AGTGGAGGGCCCTCTGCTCACTTCGAGAGAAATATCTGGTGGCTGATAGGGATACAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGGTCTTGGCTGTGCTGCCAAAGCTTACAGACCCAGAGTGTACGGGAATGTGATGGTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTTATCGACAAATCAGGGCAGACGGC 1476

RESULT 12
ADBL4381
ID ADBL4381 standard; cDNA; 1476 BP.
XX
XX ADBL4381;
AC
XX 18-DEC-2003 (first entry)
DT
XX Human prostate specific cDNA p100C ORF (minus stop codon).
DE
XX Human; ss; prostate specific cDNA; cytostatic; immunostimulant;
KW gene therapy; cell therapy; vaccine; T-cell epitope;
KW class I major histocompatibility complex allele; MHC; prostate cancer;
KW tumour; antigen presenting cell.
XX Homo sapiens.
XX US2003185830-A1.
FN
XX 02-OCT-2003.
PD
XX 12-NOV-2002; 2002US-00294025.
PF
XX 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 09-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.

PR 09-APR-1999; 99US-00288946.
 PR 13-JUL-1999; 99US-00352616.
 PR 12-NOV-1999; 99US-00439313.
 PR 18-NOV-1999; 99US-00443696.
 PR 14-JAN-2000; 2000US-00483672.
 PR 27-MAR-2000; 2000US-00536857.
 PR 09-MAY-2000; 2000US-00568100.
 PR 12-MAY-2000; 2000US-00570737.
 PR 13-JUN-2000; 2000US-00593793.
 PR 27-JUN-2000; 2000US-00605783.
 PR 09-AUG-2000; 2000US-00636215.
 PR 29-AUG-2000; 2000US-00651236.
 PR 06-SEP-2000; 2000US-00657279.
 PR 02-OCT-2000; 2000US-00679426.
 PR 10-OCT-2000; 2000US-00685166.
 PR 09-NOV-2000; 2000US-00709729.
 PR 12-JAN-2001; 2001US-00759143.
 PR 09-FEB-2001; 2001US-00780569.
 PR 09-MAY-2001; 2001US-00852311.
 PR 29-JUN-2001; 2001US-00895814.
 PR 10-DEC-2001; 2001US-00012896.
 PR 09-MAY-2002; 2002US-00144678.
 XX

(CORI-) CORIXA CORP.

Xu J, Stolk JA, Kalos MD;

WPI; 2003-756193/71.

P-PSDB; ADB14382.

New isolated polypeptide for use in a vaccine for stimulating an immune response, or for treating or diagnosis cancer, preferably prostate cancer.

Example 2; Page; 101pp; English.

The invention relates to an isolated polypeptide comprising no more than 11-542 amino acids of ADB13563 comprising a sequence ADB14487. The peptides comprise a fragment ADB13563 of that contain naturally processed T-cell epitopes for 3 class I major histocompatibility complex (MHC) alleles. ADB13563 is a polypeptide encoded by a human prostate specific cDNA, one of 648 disclosed as new. Also included are nucleic acids encoding the proteins and peptides, expression vectors, a host cell transformed with the vector, an isolated antibody for antigen binding fragment that specifically binds to the protein or peptide, detecting the presence of a cancer in a patient (comprising contacting a patient sample with a binding agent that binds to the peptides or a polypeptide appearing as ADB13563, detecting the amount of polypeptide that binds to the agent and comparing the amount of polypeptide to a predetermined cut-off value to determine the presence of cancer), a fusion protein comprising the peptides or proteins, stimulating or expanding T cells specific for a tumour protein comprising contacting T cells with the peptides or the isolated T cell population, treating prostate cancer in a patient comprising administering a composition comprising the peptides, nucleic acids, antibodies or compounds, determining the presence of a cancer in a patient and treating prostate cancer in a patient comprising incubating cluster of differentiation (CD4+) and/or CD8+ T cells isolated from a patient with the peptides or antigen presenting cells that express the peptides so that the T cells proliferate, and administering the proliferated T cells to the patient. The peptides (or an oligonucleotide that hybridises to nucleic acid encoding them), is used to detect the presence of cancer in a patient. The peptides, nucleic acids encoding, or antigen-presenting cells expressing the nucleic acid, are used to stimulate or expand T cells specific for a tumour protein. The peptides, nucleic acids, antibodies, fusion proteins, T cell populations or antigen presenting cells are used to stimulate an immune response or treat prostate cancer in a patient. The present sequence is a known cDNA showing sequence similarity to one of the disclosed human prostate specific cDNAs. Note: Except where otherwise indicated, the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030185930.

SQ Sequence 1476 BP; 361 A; 405 C; 402 G; 308 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3,21e-205 Length: 1476
 Score: 2709.00 Matches: 491
 Percent Similarity: 99.80% Conservatives: 0
 Best Local Similarity: 99.80% Mismatches: 1
 Query Match: 99.71% Indels: 0
 DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x ADB14381 (1-1476)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 1 ATGGCTTTGAACCTCAGGCTCACCACAGCTATTGGACCTTACTATGAAACCATGATAC 60
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 61 CAACCGGAAACCCCTATCCGACAGCCACCTGTGTCCCACTCTTACGAGGTGCAT 120
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 121 CCGGCTCAGTACTACCCGTCCTCCGTCAGTACGCCCGGAGGCTCTGACGAGGCT 180
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 181 TCCAAACCCGTCGTCGACGACGCCCAATCCCATCCGGACAGTGTGCACCTCAAG 240
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 Db 241 ACTAAGAAGCACTGTGTCATCACTTGGACCTTGGGACCTTCTCTGGAGCTGGCTG 300
 QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 301 GCGCTGGCTTCTCTGGAAGTTCAATGGGACAGTGTCTCAACTCTGGGATGAGTGC 360
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 361 GACTCTCAGTACTGTCATCAACCTCTAACTGTGTGTGATGGGTGTGCACACTGCC 420
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 421 GCGCGGGAGGACGAGAATCGGTGTGCTGCTCTACGATCAAACTTATCTCTCAGGTG 480
 QY 161 TyrSerSerGlyArgLysSerTyrPheProValCysGlnAspAspTyrAsnGluAsnTyr 180
 Db 481 TACTATCTCAGAGGAGTCTCTGGACCTGTGTGCCAAGACGACTGGACAGAGTAC 540
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyTle 200
 Db 541 GCGCGCGGCTCTGACGAGGACATGGGCTATAGAAATAATTTTACTCTAGCAAGAAATA 600
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 601 GTGGATGACAGCGGATCCACCACTTATGAACCTGAACCAAGTCCGCGCAATGTGAT 660
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 661 ATCTATATAAAACTGTACCAAGTGTCTCTTCAAAAGCAGTGTGTTTCTTACGC 720
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 721 TGTATAGCTGGGGGTCACTTGAACCTGAACCAAGTCCGCGCAATGTGCGCGGAG 780
 QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 781 AGCGGCTCTCCGGGGCTCTGGCGGAGTGTGACCTGCGAGGTCAGCTCCAGAAAGTCCAG 840
 QY 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGlyLys 300
 Db 841 TCGGAGGCTCCATCATCAACCCCGAGTGTGATCGTGGACAGCGGCCCTGCTGGGAAAA 900
 QY 301 ProLeuAsnAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320

```
Db 901 CCTCTTAACAAATCCATGGCATTTGACGGCATTTCGGGGATTTTGAGCAATCTTTCATG 960
Qy PheTyrGlyAlaGlyTyrGlnValGluLeuValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGAGCGCGATACCAAGTAGAAAAGTAGATTCTCATCAAAATTATGACTCCAAG 1020
Qy 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACAAATGATCTGGCGTGATGAAGCTGCAGAAGCTCTGACTTTCAACGACCTA 1080
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1081 GTGAACACAGTGTCTGCCCAACCCAGGATGATGCTGCAGCCAGACAGCTCTGCTGG 1140
Qy 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGTGGGGGGCCACGAGGAGAAAGGAGACCTCAGAAGTGTGTGAACGCTGCC 1200
Qy 401 LysValLeuLeuIleGluThrGlnAtcCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 ARGGTGCTTCTCATTTAGACACAGAGATGCAACAGCAGATATGCTATGACACCTGATC 1260
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACCAACCCATGATCTGCGGGCTTCTGCAGGGGAACGTCGATTTTCCAGGGGTGAC 1320
Qy 441 SerGlyCysProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGGCTCTGGTCACTTCGAAGAACAAATCTGGTGGCTGATAGGGGATACAAGC 1380
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGTCTGGCTGTGCAAAAGCTTACAGACCAAGGAGTGTACGGGAATGTGATGTATTC 1440
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTATTCACAATGAGGGCAGACGGC 1476
RESULT 13
AAH93943
ID AAH93943 standard; cDNA; 1479 BP.
AC AAH93943;
XX
DT 04-OCT-2001 (first entry)
XX
DE P1000C open reading frame cDNA sequence with stop codon.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytosolic; gene therapy; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US0001574.
XX
PR 14-JAN-2000; 2000US-00483672.
XX
PA (CORI-) CORIAX CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for diagnosing,
PT monitoring and treating prostate cancer in a patient and for use in
PT vaccines.
```

```
XX Claim 1; Page 538; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytosolic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated T cells
CC prepared using (I) or (II) are used to treat cancer in a patient. (I) and
CC the antibodies are also used in the detection of cancer in a patient. The
CC cancer that is diagnosed or treated is particularly prostate cancer. (I)
CC and (II) can be used in vaccines. The antibodies or (I) can be used for
CC monitoring the progression of cancer in a patient. (I) and (II) can also
CC be used to improve diagnostic and therapeutic methods for prostate
CC cancer. They can indicate the level of metastasis as well as the prostate
CC volume. AAH93357 to AAH93944 and AAH01115 to AAH01318 represent
CC polynucleotide and amino acid sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 1479 BP; 363 A; 405 C; 402 G; 309 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 3, 21e-205 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservations: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 4 Gaps: 0
```

```
US-09-615-285B-2 (1-492) x AAH93943 (1-1479)
Qy 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGTCTTGAATCAGGGTCCACCAGCTATTGGACCTTACTATGAAAACCATGGATAC 60
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCCGCACAGCCACTGTGGTCCCCTGTCTACGAGTGCAT 120
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCGCCCGTCCCGCAGTACGCCCGGAGGTCCTCAGCAGGCT 180
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCCGTCTGTGACGAGCCCAATCCCATCCGGGACAGTGTGACCTCAAG 240
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAAGCACGTGTGCATCACCTTGACCTGGGACCTTCTCTGCGGAGCTCGGCTG 300
Qy 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCCGCTGGCCTACTCTGGAGTTCTATGGGACAGAGTGTCTCAACTCTGGGATAGAGTGC 360
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCTCAGGTACCTGTGCATCAACCCCTCTAACTGGTGTGATGGGTGTCACTGCCCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuVal 160
Db 421 GCGCGGAGGACGAGAATCGGTGTCTCCCTCTACCGATCAAACTTCTATCTTCAGGTG 480
Qy 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGGAAGTCTCGGACCTCTGTGTCGAAGACGACTGGAACGAGAATAC 540
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GCGCGGCGGCTGCAGGAGACATGGGCTATAAGATAAATTTTACTCTAGCCACGAATA 600
Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACGACTTTATGAACCTTTATGAACAGACAGTGGCCGCAATCTCGAT 660
```

```
QY 221 lletYrLysLysLeuYrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAACCTGACACAGTGATCGCTGTTCTTCAAAGCAGTGCTTCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 721 TGTATAGCCTCGGGGGTCACTTGAACTCAGCCGACAGCAGATGTTGGCGGGAG 780
QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGGCTCCCGGGGCTGGCCCTGGCAGTGTCAGCTGACGTCACGTCAGAACGTCACGTC 840
QY 281 CysGlyGlySerIleIleThrProGlnTrpIleValThrAlaAlaHisCysValGluLys 300
Db 841 TCGCGAGCTCCATCATCACCCCGAGTGGATCGTGCAGCGCCCACTGCGTGGAATA 900
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTACATCATCGATGTCATGGACGCACTTGGGGGATTTGAGACATCTTTTCATG 960
QY 321 PheYrGlyAlaGlyYrGlnValGluLysValIleSerHisProAsnTyAspSerLys 340
Db 961 TTCTATGAGCGGATACCAAGTAGAAAAGTGATTTCTCATCCAAATATGACTCCAAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGATTCGCTGATGAAGCTGCAGAAAGCTCTGACTTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1081 GTGAACACAGTGTCCTGCCAACCCAGCGCATGATGCTGCAGCCAGAACACCTCTGCTGG 1140
QY 381 IleSerGlyTrpGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCGGGTGGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTGCTGAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgYrValYrAspAsnLeuIle 420
Db 1201 AAGTGCTTCTCATTTGACACACAGATGTCACAGCATATGTCATGACAACTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACACGACCATGATCTGTGCGGCTTCTCTGAGGGGAAACGTCGATTTCTGACGGGTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGAAGGCGCTCTGGTCACCTTCGAAGAACAAATATCTGGTGCTGATAGGGGATACAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyArgProGlyValYrGlyAsnValMetValPhe 480
Db 1381 TGGGGTCTGTGCTGTGCCAAAGCTTACAGACAGGAGTGTCAGGGAATGTGATGTTATTC 1440
QY 481 ThrAspTrpIleTyArgGlnMetArgAlaAspGly 492
Db 1441 ACGGATGATTTATCGACAAATGAGGCGACAGCGC 1476
RESULT 14
AAS64179
ID AAS64179 standard; cDNA; 1479 BP.
XX AC AAS64179;
XX DT 29-JAN-2002 (first entry)
XX DE Human prostate cDNA sequence #598.
XX KW Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.
XX OS Homo sapiens.
XX PN WO200173032-A2.
XX
```

```
PD 04-OCT-2001.
XX 27-MAR-2001; 2001WO-US009919.
XX 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 10-OCT-2000; 2000US-00679426.
PR 18-OCT-2000; 2000US-00685166.
PR 03-NOV-2000; 2000US-00709729.
XX (CORI-) CORIXA CORP.
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GS, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX WPI; 2001-639232/73.
DR P-PSDB; AAU69960.
XX New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer.
XX Claim 1; Page 571-572; 579pp; English.
XX The invention relates to isolated prostate-specific polynucleotides,
CC polypeptides, fusion proteins of the polypeptides, antibodies raised
CC against the polypeptides (or antigenic epitopes derived from them) and
CC antigen-presenting cells expressing the polypeptides. The antibodies are
CC useful for detecting the presence of cancer, especially prostate cancer.
CC The polypeptides, polynucleotides and the antigen-presenting cells are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein, and for inhibiting the development of cancer especially prostate
CC cancer. Compositions comprising the polynucleotide and/or polypeptide are
CC useful for stimulating an immune response, and for treating cancer. The
CC oligonucleotide is useful for detecting cancer. The present sequence is a
CC prostate specific polynucleotide of the invention
XX SQ Sequence 1479 BP; 363 A; 405 C; 402 G; 309 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 3-21e-205 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 4 Gaps: 0
US-09-615-285B-2 (1-492) x AAS64179 (1-1479)
QY 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyTyTyGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAACCTCAGGGTCACCCAGCTATTGGACCTTACTATGAAACCATGATAC 60
QY 21 GlnProGluAsnProTyTyProAlaGlnProThrValValProThrValTyTyGluValHis 40
Db 61 CAACCGGAAACCCCTATCCGCACAGCCACTGTGGTCCCTCTGTACGAGGTGCAT 120
QY 41 ProAlaGlnTyTyTyProSerProValProGlnTyTyAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCCTCCCGCCAGTACGCGCCGAGGGTCTTCGACGAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCACACCCCGTCTGTGCAGCAGAGCCCAATCCCATCCGGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
```

Db 241 ACTAAGAAAGCACTGTGCATCACTTGACCTGGGACCTTCTCTGTGGAGCTGGCTG 300
Qy 101 AlaalaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCGCTGGCTACTCTCGAAGTTTCATGGGAGCAAGTCTCCAACTCTGGATAGAGTGC 360
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGGTACTCTGCATCACTCCCTCTAACTGGTGTGATGGCGTGCACACTGCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTrpGlyProAsnPhelLeuGlnVal 160
Db 421 GCGGGGAGGACGAGAACTCGGTGTGTTCGGCTCTACGGATCAAACTTCATCTTCAGGTG 480
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 481 TACTCTCTCAGAGAGTCTTGCACCTCTGTGCCAAGCAGCTGGAACGAGACTAC 540
Qy 181 GlyArgAlaAlaCysAAspMetGlyTyrIleAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GCGGGGCGGCTCGCAGGACATGGGTATAAGAAATAATTTTACTCTAGCCCAAGAAATA 600
Qy 201 ValAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAAAGTCCGCGCAATGTGAT 660
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAACCTGTACCAGATGCTGCTTCTTCAAAGCAGTGGTTCTTTACGC 720
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 721 TGTATAGCTCGGGGCTCACTTGAACCTCAAGCGCCAGAGCGAGATGTGGCGCGGAG 780
Qy 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGGCTCCGGGGGCTGGCTTGGAGGTGAGCTGCAGCTCCAGAACGTCACGCTG 840
Qy 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 841 TCGGAGGCTCATCATCACCCCGAGTGGATCGTGACAGCCGCCACTCGCTGGGAAAA 900
Qy 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACATTCATGGCATTTGACGGCATTTGCGGGGATTTTGACACAATCTTTCATG 960
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATTTATGACTCCAA 1020
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGATTCGCTGATGAGCTGCAGAGCTCTGACTTTCAACGACTTA 1080
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1081 GTGAACACAGTGTCTCCCAACCCAGGCATGATGTGCAGCCAGAACAGCTCTGTGTTG 1140
Qy 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGGTGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTGTGAACGCTGCC 1200
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGGTGCTTCTCATTTGACACACAGATGCAACAGCAGATGCTATGACAACTGATC 1260
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACCAAGCATGATCTGTGGGGCTTCTGTGAGGGAAAGCTCGATTCTTTCAGGGTGAC 1320
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGCGCTCTGGTCACTTCGAAAGAACATATCTGGTGGCTGTAGGGGATACAGC 1380

Qy 461 TrpGlySerGlyCysAlaLysAlaLysValTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGTTCTGGCTGTGCCAAAGCTTACAGACAGAGGTGTACGGGAATGTGATGTATTC 1440
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTTCGACAAATGAGGCGACAGCGC 1476

RESULT 15

ACA59987
ID ACA59987 standard; cDNA; 1479 BP.

XX ACA59987;

XX 10-JUN-2003 (first entry)

XX Prostate cancer therapy associated cDNA #668.

XX Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
KW immunogen; cancer; prostate specific antigen; PSA;
KW prostatic acid phosphatase; PAP; prostate specific membrane antigen;
KW PSMA; gene; ss.

XX Homo sapiens.

XX US2002192763-A1.

XX 19-DEC-2002.

XX 29-JUN-2001; 2001US-00895793.

XX 04-OCT-1999; 99US-0157455P.

XX 04-OCT-2000; 2000US-00679272.

XX 28-MAR-2001; 2001US-00822827.

XX (XUJ/) XU J.

XX (DILL/) DILLON D C.

XX (MITC/) MITCHAM J L.

XX (HARL/) HARLOCKER S L.

XX (JIAN/) JIANG Y.

XX (KALO/) KALOS M D.

XX (FANG/) FANGER G R.

XX (RETT/) RETTER M W.

XX (STOL/) STOLK J A.

XX (DAYC/) DAY C H.

XX (VEDV/) VEDVICK T S.

XX (CARL/) CARTER D.

XX (LISX/) LI S X.

XX (WANG/) WANG A.

XX (SKEI/) SKEIKY Y A W.

XX (HEPL/) HEPLER W T.

XX (HEND/) HENDERSON R A.

XX (HURA/) HURAL J.

XX (MCNE/) MCNEILL P D.

XX (HOUG/) HOUGHTON R L.

XX (DBAS/) Y DE BASSOLS C V.

XX (FOYT/) FOY T M.

XX Xu J. Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Y De Bassols C, Hural J;
PI McNeill PD, Houghton RL, Y De Bassols C, Foy TM;
XX WPI; 2001-245062/25.
XX Prostate specific protein and its encoding polynucleotide, useful for the
PT treatment and diagnosis of prostate cancer.
XX Example 2; SEQ ID NO 930; 85pp; English.
XX The invention describes a fusion protein comprising at least one amino
CC acid sequence of immunogenic portions of any of the 3 sequences not

CC defined in the specification, or sequences having at least 70 or 90 %
CC sequence identity to any one of the 35 sequences defined in the USPTO web
CC site, which is encoded by any of the 4 nucleotide sequences not defined
CC in the specification. The fusion protein, composition and methods are
CC useful for diagnosing, preventing and/or treating cancer, particularly
CC prostate cancer. The proteins are useful as markers to indicate the
CC presence or absence of cancer. This sequence represents a prostate cancer
CC therapy associated cDNA. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20020192763
XX
SQ Sequence 1479 BP; 363 A; 405 C; 402 G; 309 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,21e-205 Length: 1479
Score: 2709.00 Matches: 451
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 5 Gaps: 0

US-09-615-285B-2 (1-492) x ACA59987 (1-1479)

QY	1	MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr	20
DB	1	ATGGCTTTGAACCTCAGGGTCAACACAGCTATTGGACCTTACTATGAAACCATGATGATAC	60
QY	21	GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis	40
DB	61	CAACCGGAAACCCCTATCCGGCAGAGCCACTGTGTCCCTCCACTGTCTACGAGGTGCAT	120
QY	41	ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla	60
DB	121	CGGGCTCAGTACTACCGTCCCGCTGCCAGTACGCCCGCCGAGGGTCTCGACGAGGCT	180
QY	61	SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys	80
DB	181	TCCAAACCCCGTGTCTGACGAGCCAAATCCCATCCCGGACAGTGTGACCTCAAG	240
QY	81	ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu	100
DB	241	ACTAAGAAAGCACTGTGCATCCTTGACCTGGGGACCTTCTCTGGGGAGCTGGCTG	300
QY	101	AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys	120
DB	301	GCAGCTGGCCCTACTCTGGAAGTTTCATGGGCAGCAAGTCTCCAACCTCTGGGATAGTGC	360
QY	121	AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro	140
DB	361	GACTCTCAGGTACCTGCATCAACCCCTCTAACCTGTGTGTGGGTGTGTCACACTGCC	420
QY	141	GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal	160
DB	421	GGCGGGGAGGACGAGAATCGGTGTGTGTGCTCTACGGATCAAACTTCTATCTCAGGTG	480
QY	161	TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspTrpAsnGluAsnTyr	180
DB	481	TACTCATCTCAGAGGAAGTCTGGCCCTGTGTGCCAAGACACTGGGAACGAGAACTAC	540
QY	181	GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle	200
DB	541	GGCGGGGCGCTTCAGGGACATGGGCTATAGAAATATTTTACTCTAGCCAAAGGAATA	600
QY	201	ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp	220
DB	601	GTGGATGACAGCGGATCCACAGCTTTATGAAACTGAACACAGTCCCGGCAATGTCGAT	660
QY	221	IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg	240
DB	661	ATCTATTAATAAACTGTACCACTGATGCTGTCTTCTTCAAAAGCAGTGTCTTTTACGC	720
QY	241	CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu	260

Search completed: June 1, 2004, 14:59:22
Job time : 668 secs

DB	721	TGATAGCTCGCGGGTCAACTTGAACCTCAAGCCGCGAGAGGATTGTGGCGCGAG	780
QY	261	SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal	280
DB	781	AGCCCGTCCCGGGGGCTGGCCCTGCGAGTCAAGCTGCGACGTCAGAACGTCACGTCG	840
QY	281	CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys	300
DB	841	TGCGAGGCTCCATCATCAATCCCGGAGTGGATCGTACAGCGCCCACTGCGTGAAAAA	900
QY	301	ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet	320
DB	901	CCTCTTAACAATCATGGCATTTGGACGGCATTTTGGGGGATTTTGAGACAACTTTTCATG	960
QY	321	PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys	340
DB	961	TTCTATGGAGCCGGATACCAAGTAGAAAAAGTAGTTTCTCATCAAAATTAATGATCTCCAG	1020
QY	341	ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu	360
DB	1021	ACCAAGAACAAATGACATTTGGCTGATGAGCTGAGAGCTGAGAAAGCTCTGACTTTCAACGACCTA	1080
QY	361	ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp	380
DB	1081	GTGAACCAAGTGTCTGCCCAACCCAGGATGATGCTGCGACGACGACAGCTCTGCTGG	1140
QY	381	IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla	400
DB	1141	ATTTCCGGTGGGGGGCCACCGAGGAGAAAGGGAAGACCTCAGAAAGTGTCTGAACGCTGCC	1200
QY	401	LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle	420
DB	1201	AAGTGTCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGTCTATGACCACTGATC	1260
QY	421	ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp	440
DB	1261	ACACAGCCATGATCTGTGCGGGTCTCTGACGGGAACTCGATTTCTTGCAGGGGTGAC	1320
QY	441	SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer	460
DB	1321	AGTGGAGGGCTCTGTGTCATCTCGAAGAACAAATATCTGGTGGCTGATAGGGGATACAGC	1380
QY	461	TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe	480
DB	1381	TGGGTTCTGGCTGTGCCAAAGCTTACAGACCAGAGTGTACGGGAATGTGATGTTATTC	1440
QY	481	ThrAspTrpIleTyrArgGlnMetArgAlaAspGly	492
DB	1441	ACGGACTGGATTTTATCGACAAATGAGGGCAGACGGC	1476

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 1, 2004; 14:39:10 ; Search time 5848 Seconds
(without alignments)
3646.507 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 2717
Sequence: 1 MALNSGPPAIGPYENHGY.....VGNVNVFTDWIYQMRADG 492

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09615285/runat_01062004_132133_19946/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTWT=ptc -NORM=ext -HRAPIZE=500 -MAXLEN=2000000000
-USER=US09615285 @CGN_1_1_3731 @runat_01062004_132133_19946 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rtd.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2717	100.0	1479	6	AR25345	Sequence
2	2717	100.0	1738	6	AX395316	Sequence
3	2717	100.0	1738	6	BD218220	Tumor ant
4	2717	100.0	1740	9	AF270487	BC051839 Homo sapi
5	2717	100.0	3195	9	BC051839	Homo sapi
6	2714	99.9	1479	9	AF123453	Homo sapi
7	2711	99.8	3226	9	AF329454	Homo sapi
8	2709	99.7	1476	6	AX201158	Sequence
9	2709	99.7	1476	6	AX267957	Sequence
10	2709	99.7	1479	6	AX201157	Sequence
11	2709	99.7	1479	6	AX267956	Sequence
12	2709	99.7	3245	6	AX201156	Sequence
13	2709	99.7	3245	6	AX267955	Sequence
14	2704	99.5	3244	6	BD244778	Specific
15	2704	99.5	3443	6	BD228306	Method of
16	2696	99.2	2479	6	BD235838	A novel m
17	2696	99.2	2479	6	AR235372	Sequence
18	2696	99.2	2479	6	AR405834	Sequence
19	2696	99.2	2479	6	AX201151	Sequence
20	2696	99.2	2479	6	AX207965	Sequence
21	2696	99.2	2479	6	AX267920	Sequence
22	2696	99.2	2479	6	AX306771	Sequence
23	2696	99.2	2479	6	AX395318	Sequence
24	2696	99.2	2479	6	BD218221	Tumor ant
25	2696	99.2	2479	9	HSU75329	Human serin
26	2696	99.2	3966	6	AX041973	Sequence
27	2121	78.1	3175	10	BC054348	Mus muscu
28	2119	78.0	3180	10	BC061712	Rattus no
29	2117	77.9	1735	10	AF243500	Mus muscu
30	2117	77.9	3014	10	BC038393	Mus muscu
31	2116	77.9	1753	10	AF113596	Mus muscu
32	2107	77.5	1755	10	AF199362	Mus muscu
33	2083	76.7	1473	10	AB073550	Rattus no
34	1569	57.7	1077	6	AR194808	Sequence
35	1484	54.6	1001	6	AR270400	Sequence
36	1165	42.9	683	6	AR405835	Sequence
37	1165	42.9	683	6	AX201123	Sequence
38	1165	42.9	683	6	AX267922	Sequence
39	1009	37.1	3507	5	AB038497	Xenopus 1
40	894	32.9	2948	6	AX354825	Sequence
41	891.5	32.8	2412	6	AX375996	Sequence
42	891.5	32.8	2412	9	AY358458	Homo sapi
43	891.5	32.8	2418	6	BD173589	Novel ser
44	885	32.6	1359	6	AX354827	Sequence
45	885	32.6	1359	6	BD173588	Novel ser

ALIGNMENTS

RESULT 1

AR225345 AR225345 1479 bp DNA linear PAT 20-DEC-2002
LOCUS Sequence 1 from patent US 644419.
DEFINITION AR225345
ACCESSION AR225345
VERSION AR225345.1 GI:27263288
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1479)
AUTHORS Wong,A.K.C., Tavtigian,S.V. and Teng,D.H.F.
TITLE TWPRSS2 is a tumor suppressor
JOURNAL Patent: US 644419-A 1 03-SEP-2002;
FEATURES
Location/Qualifiers
1..1479
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3,14e-197 Length: 1479
Score: 2717.00 Matches: 492
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-615-285B-2 (1-492) x AR225345 (1-1479)
QY 1 MetAlaLeuAenSerGlySerProAlaIleGlyProTyrTyrGluAenHisGlyTyr 20
Db 1 ATGGCTTTGAATCAGGCTCACCACAGCTATTGGACCTTACTATGAAACCACTGGATAC 60
QY 21 GlnProGluAenProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCCGCACAGCCCACTGTGTCCCACTGTCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCGCTCCCGTCCCGTCCCGTACGACGCGGCTCTCTGACGAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCCGTGTGTGACGACGACCCCAATCCCAATCCCGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
Db 241 ACTAAGAAAGACACTGTGCATCACCTTGACCTCGGGACCTTCTCTGGGAGCTCGCGTG 300
QY 101 AlaAlaGlyLeuLeuTyrPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCCGCTGCCCTACTCTGGAAGTTTCATGGGACAGAGTCTCCAACTCTGGGATAGATGC 360
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGGTACCTGTGCATCAACCCCTTAACCTGTGTGTGTGTGTGTGTGTGTGTGT 420
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 421 GGCGGGGAGGACGAGATCGGT 480
QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 481 TACTCATCTTCAGAGGAAGTCTCGGACCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 541 GGCGGGGCGGCTGAGGACATGGGCTATAGATAATATTTTACTCTTAGCCCAAGGATA 600
QY 201 ValAspAspSerCysThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACAGCTTTATGAAACTGAAACACAGTGCCTGGCAATGTGCAT 660

QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATATAAAACCTGTACCACAGTGTGCTGTCTTTCAAAAGCAGTGGTTCCTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgInSerArgIleValGlyGlyGlu 260
Db 721 TGTATAGCTCGGGGCTCAACTTGAACCTCAAGCCCGCAGACAGCATCTGTGGGCGGAG 780
QY 261 SerAlaLeuProGlyValAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGGCTCCCGGGGGCTGGCCCTGGCAGGTGACCTGACCTGCCAGCTCCAGACGTCCACGTG 840
QY 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
Db 841 TGGGAGGCTCCATCATCACCCCGGAGTGGATCTGACAGCCGCCCACTGTGGTGGAAAAA 900
QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACAATCCATGGCATTTGGAGCATTTGGGGGATTTTGAGACATCTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGAGCGGGATACCAAGTAGAAAAAGTATTTCTATCCAAATATATGACTCCAAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACAAATGACATTCGCTGATGAAGCTGAGAGAGCTCTGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1081 GTGAACACAGTGTGTCTGCCAACCCAGCATGATGCTGCAGCCAGAAACAGCTCTGTGG 1140
QY 381 IleSerGlyTyrGlyValThrClnLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGGTGGGGGCCACCCAGGAGAAAGGAAAGACCTCAGAAGTGTCTGNACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1201 AAGTGTCTTCTATTGAGACACAGAGATGCAACAGACATATGCTATGACAAACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACCCAGCATATCTGTGCCGCTTCTGCGGGGAAAGCTGATTTCTGCCAGGGTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
Db 1321 AGTGAGGGCCCTCTGTCTTCTGAGAACATATCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGGTCTGT 1440
QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGGACTGGATTATCGACAAATGAGGGCAGACGGC 1476
RESULT 2
AX395316 1738 bp DNA linear PAT 18-MAY-2002
LOCUS Sequence 1 from Patent WO0204953.
DEFINITION AX395316
ACCESSION AX395316.1 GI:21066313
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Safran,D., Raitano,A.B., Hubert,R.S., Jakobovits,A., Faris,M. and
Challita-Eid,P.M.
TITLE Novel tumor antigen useful in diagnosis and therapy of bladder,
ovary, lung and kidney cancers
JOURNAL Patent: WO 0204953-A 1 17-JAN-2002;

Agensys, Inc. (US)
 Location/Qualifiers
 1. 1738
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
 Pred. No.: 3.81e-197 Length: 1738
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x AX395316 (1-1738)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 DB 112 ATGGCTTTGAACCTAGGGTCAACCCAGCTATTGGACCTTACTATGAACCATGGATAC 171
 QY 21 GlnProGluAunProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 DB 172 CAACCGAAACCCCTATCCCGCACAGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 231
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 DB 232 CCGGCTCAGTACTACCGCTCCCGGTGCCAGTACGCCCGGAGGCTTCTGACACAGCT 291
 QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 DB 292 TCCAAACCCGCTGCTGCACGAGCCCAATCCCAATCCCGACAGTGTGCACCTCAAG 351
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 DB 352 ACTAAGAAAGACACTGTGCATCACTTACCTGAGGACCTTCTCCCTGGGAGCTGCCGTG 411
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 DB 412 CCGGCTGCCCTACTCTGAGTTCATGGGACGAGTCTCAACTCTGGGATAGATGC 471
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 DB 472 GACTCCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGGTGCACATGCCCC 531
 QY 141 GlyGlyGluAapGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 DB 532 GCGGGGAGGAGGAGAACTGGTGTTCCTCTACGACCAACCACTTCACTCTCAGGTG 591
 QY 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
 DB 592 TACTCATCTCAGAGGAAGTCTCTGGCACTGTGTGCCAAGACGACTGGAACGAGAAGTAC 651
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
 DB 652 GCGGGGGGGGCTCGAGGGACATGGCTATAGAAATAATTTTACTCTAGCCAAAGGAATA 711
 QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 DB 712 GTGGATGACGCGGATCCACCGCTTTATGAACCTGAACACAGTGGCGGCAATGTGCAT 771
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 DB 772 ATCTATAAAAACTTACCACAGTATGCTGTTCTTCAAAAAGCAGTGGTTCTTTACGC 831
 QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 DB 832 TGATAGCTGCGGGGTCAACTTGAACCTCAAGCCCGCAGAGCAGGATTTGTGGCGGCGAG 891
 QY 261 SerAlaLeuProGlyValatTrpProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 DB 892 ASCGCGCTCCCGGGGCTGGCCCTGGCAGGTGACCTGCAGTCCAGAAAGTCCAGCTG 951

QY 281 CysGlyGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
 DB 952 TCGGAGGCTCCATCATCCCCCGAGTGGATCGTGACAGCGCCACCTGCGTGGAAAAA 1011
 QY 301 ProLeuAsnAsnProTyrPheIleThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 DB 1012 CTTCTTAACAATCCATGGCATTTGGACGGCATTTGGGGGATTTTGAGACAAATCTTTCATG 1071
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 DB 1072 TTCTATGGAGCGGATACCAAGTAGAAAAGATTTCTCATCCAAATATGACTCCAAAG 1131
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 DB 1132 ACCAAGAACAATGACATTTGGCTGATGAGCTGCAGAGCCCTCTGACTTTCAACGACCTA 1191
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
 DB 1192 GTGAACACAGTGTCTGCCCAACCCAGGCATGATGCTGCAGCCAGACAGCTCTGCTGG 1251
 QY 381 IleSerGlyTyrGlyValaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 DB 1252 ATTTCGGGTGGGGGCCACCCAGGAGAAAGGAGAACCTCAGAGTGTCTGAACGCTGCC 1311
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 DB 1312 AAGGTCTTCTCATTTGAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1371
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 DB 1372 ACACCAAGCATGATCTGTCCGCTTCTTCAGAGGGAAGCTGATTTCTTCAGAGGTGAC 1431
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
 DB 1432 AGTGGAGGCTCTGTCTCACTTCGAGAGAACATATCTGTGCTGATAGGGGATACAAGC 1491
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 DB 1492 TGGGGTCTCTGGCTGTGCCAAGCTTACAGACAGGAGTGTACGGGAATGTGATGTTTC 1551
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 DB 1552 ACGGACTGATTTATGCAAAATGAGGGGACAGCGC 1587

RESULT 3
 BD218220 1738 bp DNA linear PAT 17-JUN-2003
 LOCUS
 DEFINITION
 Tumor antigen useful in diagnosis and therapy of prostate and colon cancer.

BD218220
 VERSION BD218220.1 GI:33027990
 KEYWORDS JP 2002517185-A/1.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1738)
 AFAR,D.E., Hubert,R.S., Leong,K., Raitano,A.B., Saffran,D.C. and Mitchell,S.C.
 Tumor antigen useful in diagnosis and therapy of prostate and colon cancer
 Patent: JP 2002517185-A 1 18-JUN-2002;
 UROGENESIS INC
 OS Homo sapiens (human)
 PN JP 2002517185-A/1
 PD 18-JUN-2002
 PF 01-JUN-1999 JP 2000552152
 PR 01-JUN-1998 US 60/087598,29-JUN-1998 US 60/091474 PR
 14-APR-1999 US 60/129521
 PI DANIEL E AFAR, RENE S HUBERT, KAHAN LEONG, ARTHUR B RAITANO PI
 DOUGLAS C SAFFRAN,
 PI STEPHEN CHAPPELL MITCHELL

PC C12N9/64,A61K38/00,A61K39/395,A61K39/395,A61P13/08,A61P35/00,
 PC C07K14/47,
 PC C07K16/40,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N15/09 PC
 C12P21/08,C12Q1/68,
 PC G01N33/50,G01N33/566,G01N33/574,A61K37/02,C12N5/00,C12N15/00
 CC Tumor antigen useful in diagnosis and therapy of prostate and
 CC colon cancer
 FH Key Location/Qualifiers
 FT CDS (112)..(1588).

FEATURES

source
 1. .1738
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 3,81e-197 Length: 1738
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservations: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x BD218220 (1-1738)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 112 ATGGCTTTGAACCTCAGGGTACCACAGCTATTGGACCTTACTATGAAACCATGATAC 171
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 172 CAACCGGAAACCCCTATCCCGACAGCCACCTGTGTGCCCTCTCTACGAGGTGCAT 231
 QY 41 ProLacGlnTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 232 CCGGCTCAGTACCTACCGCTCCCGTCCCGACGACCCCGAGGGTCCCGACGAGGT 291
 QY 61 SerAsnProValValCysThrGlnProLysProSerGlyThrValCysThrSerLys 80
 Db 292 TCCAAACCCGTCGTCACGACGACCAATCCCATCCGACGAGTGTGCACCTCAAAG 351
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 Db 352 ACTAAGAAAGCAGTGTGCATCACCTTGACCTGGGGACCTCTCTCGGGAGGTGCGGTG 411
 QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 412 GCCGCTGGCCTACTCTGGAAGTTCATGGGACGCAAGTCTCCAACTCTGGGATAGAGTGC 471
 QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 472 GACTCCTCAGGTACTCTGCATCAACCCCTCTAACTGGTGTGATGGCGGTGCACACTGCCCC 531
 QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 532 GCGCGGAGACGAGATCGGTGTTCGCTCTACGACCAACCTCATCTCTTCAGGTG 591
 QY 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
 Db 592 TACTCATCTCAGAGAAAGTCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGAACTAC 651
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 652 GGGCGGGCGGCTTCAGGGACATGGGCTATAGAAATAATTTTACTTACTACCCAGGATA 711
 QY 201 ValAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 712 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACCAAGTCGCGCAATGTGAT 771
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 772 ATCTATAAAAACTGTACACAGTATGCTCTTCTTCAAAAGCGTGTCTTCTTACGC 831

QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
 Db 832 TGTATAGCTCGGGGTCAACTTGAATCAAGCCGACAGCAGGATTGTGGGGCGAG 891
 QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 892 AGCGGCTCCCGGGGCTGCGCTGGCAGGTACGCTGCACGTCCAGACGTCCACGTG 951
 QY 281 CysGlyGlySerIleIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
 Db 952 TCGGAGGCTCCATCATCACCCCGAGTGGATCTGACAGCGGCCCTCGTGGAAAAA 1011
 QY 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 1012 CCTCTTAACAATCCATGCGCATTTGACGSCATTTGCGGGGATTTTCAGACAATCTTTCATG 1071
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 1072 TTCTATGAGCGGATACCAAGTAGAAAAAGTATTCTCTATCCAAATATTAGCTCCAG 1131
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1132 ACCAAGAACAATGACATTGCGCTGATGAGGTGAGAGCCTCTGACTTTTCAAGACCTA 1191
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 Db 1192 CTGAAACCCAGTGTGTGCCCCAACCCAGCATGATGCTGCAGCCAGAACAGCTCTGTGG 1251
 QY 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
 Db 1252 ATTTCCGGGTGGGGGCGCCCGAGAGAAAGGAGACCTCGAAGTGTCTGACGCTGCC 1311
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1312 AAGTGTCTCTCATTTAGACACAGATGCAACAGCAGATATGCTATGACCAACCTGATC 1371
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1372 ACACAGGACCATGATCTGTGCGGCTTCTCGAGGGAAACGTCGATCTTTCGCGGTGAC 1431
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrPheLeuIleGlyAspThrSer 460
 Db 1432 AGTGAGGGCTCTGCTGCTCAATTCGAAGAACAAATATCTGGTGGCTGATAGGGGATACAAGC 1491
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1492 TGGGTTCTGCTGTGCCAAGCTTACAGACAGAGGTACGGAGTGTACGGATGTGTATTC 1551
 QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
 Db 1552 ACGGACTGGATTATCGACAAATGAGGGCAGACGSC 1587

RESULT 4

AF270487 1740 bp mRNA linear PRI 04-APR-2001
 LOCUS
 DEFINITION Homo sapiens androgen-regulated serine protease TMPRSS2 precursor (TMPRSS2) mRNA, complete cds.
 ACCESSION AF270487
 VERSION AF270487.1 GI:13540003
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1740)
 Afar,D.E., Vivianco,I., Hubert,R.S., Kuo,J., Chen,E., Saffran,D.C., Raitano,A.B. and Jakobovits,A.
 Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia
 Cancer Res. 61 (4), 1686-1692 (2001)
 21139112
 11245484
 PUBMED

REFERENCE 2 (bases 1 to 1740)
 AUTHORS Mitchell, S.C., Hubert, R.S. and Afar, D.E.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAY-2000) UroGenesys, Inc., 1701 Colorado Ave., Santa Monica, CA 90404, USA

FEATURES
 source
 1..1740
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 1..1740
 /gene="TMPRSS2"
 114..1592
 /gene="TMPRSS2"
 /note="expressed in prostate and prostate cancer; differs from product encoded by the sequence presented in GenBank Accession Number U75329"
 /codon_start=1
 /product="androgen-regulated serine protease TMPRSS2 precursor"
 /protein_id="AAK29280.1"
 /db_xref="GI:13540004"
 /translation="MALNSGPPAIGPYENHGYQENPYPAQPTVPTVVEHPAQY YSPVQAPVRLTQASNPVCTQPKSPGTVCTKTKKALCITLTGTFVUGAALAA GLLKFMGSKCSNGIEDCSGTCINPNCWCDGVSHCPGGEDENRCLVLYPNFLLQV YSSQRKSHVPCQDWNENYGRAARDGMGYKNFYSSQGIYVDDSGSISFMKLTNSAGN VHYKXLYHSDASKSAVVSURCIACGNLNSROSRIYVGSALPGAPWQVSLRVQ NVHYGGGSIITPEWTVTAHCKVEKPELNNPWHWTFAGILRQSFYFGAPYQVEVISH PNYDSKTNKNDLIALKLOKPLTFNDLVRPVLCPNPNMLOPEQLCMTSGWGATEEKGK TSEVLNAKKNLIIETQRNSRYVDNLITPAMICAGFLQGNVDSCQDGSGLPVTSKN NIWLIGDTSWGSQCAKAPRGVYGVNVMVFTDWIYROWRADG"
 879..1589
 /gene="TMPRSS2"
 /product="androgen-regulated serine protease TMPRSS2"

mat_peptide
 879..1589

ORIGIN
 Alignment Scores:
 Pred. No.: 3,82e-197 Length: 1740
 Score: 2717.00 Matches: 492
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-615-285B-2 (1-492) x AF270487 (1-1740)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 114 ATGGCTTTGACCTACGGGTCACCCAGCTATTGGACCTTACTATGAAACACCTGGATAC 173
 QY 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
 Db 174 CRACCGGAAACCCCTATCCGACAGCCCACTGTGGTCCCACTGTCTACGAGGTGCAT 233
 QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 234 CCGGCTCAGTACTACCGCTCCCGCTGCCAGTACCGCCCGCCGAGGCTCTGACGAGGCT 293
 QY 61 SerAsnProValValCysThrGlnProTyrSerProSerGlyThrValCysThrSerLys 80
 Db 294 TCCAAACCCGCTGCTGACGAGCCCAATCCCAATCCGAGACAGTGTGACCTCAAG 353
 QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
 Db 354 ACTAAGAAAGCAGTGTGCATCACCTTGACCCCTGGGACCTTCTCGTGGAGCTCGGCTG 413
 QY 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 414 GCGGCTGGGCTACTCTGGAAGTTATGGGACAGAGTGTCTCAACTCTGGGATAGAGTGC 473
 QY 121 AspSerSerGlyThrCysIleLeuProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 474 GACTCTCAGGTACCTGCACTCAACCCCTTACTGCTGATGGGCTGTCACTGCCCC 533

QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 534 GCGGGGAGGACGAGAATCGTGTGTGGCTCTACGACCAAACTTCATCTTCAGGTG 593
 QY 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspSerTyrPheGlnAsnTyr 180
 Db 594 TACTCATCTCAGAGGAAGTCTTGGCACCTGTGTGCAAGACGACTGGAAACGAGAACTAC 653
 QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 654 GGGCGGGCGCTGCAGGGACATGGGCTATAAGAAATAATTTTACTCTAGCCAAGGAATA 713
 QY 201 ValAspSerSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 714 GTGGATGACAGCGGATCCACCGCTTATGAAACTGAACAAGTCCGCGCAATGTTCGAT 773
 QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 774 ATCTATAAAAACCTGTACCACTGATGCTCTTCTTCAAAAGCAGGTGTTTCTTTACGC 833
 QY 241 CysIleAlaCysGlyValLeuAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
 Db 834 TGTATAGCTCGCGGGTCAACTTGAACCTCAAGCGCCGACGAGGATTTGGCGCGCGAG 893
 QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
 Db 894 AGCGCGCTCCCGGGGCTGGCCCTGGCAGTCAGCTGCACCTCCAGACGTCACGCTG 953
 QY 281 CysGlyGlySerIleThrProGluTyrPheValThrAlaAlaHisCysValGluLys 300
 Db 954 TCGGAGGCTCCATCATCACTCCCGGCTGATGCTGCACAGCGCCCACTCGTGGAAAAA 1013
 QY 301 ProLeuAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
 Db 1014 CCTCTTAACAATCCATGGCATTTGACGSCATTTTGGGGGATTTTGAGACAACTCTTTCATG 1073
 QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 1074 TTCTATGAGCGCGATACCAAGTAGAAAAAGTAGTTTCTCATCCAAATATGATCCAAAG 1133
 QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1134 ACCAAGAACCAATGACATTCGCTGATGAGCTGAGAGCTCTGACTTTCAACGACCTA 1193
 QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 Db 1194 GTGAAACCACTGTCTGCTGCCAACCCAGGCACTGATGCTGCAGCCAGAACAGCTGTGGT 1253
 QY 381 IleSerGlyTyrGlyAlaThrGluGlyGlyThrSerGluValLeuAsnAlaAla 400
 Db 1254 ATTTCCGGGTGGGGGCGCCACCGAGGAGAAAGGAAAGACCTCAAGAGTCTGAACGCTGCC 1313
 QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
 Db 1314 AAGGTGTTCTCTATTGAGACACAGAGATGCAACAGCAGATATGCTATGACAACTGATC 1373
 QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1374 ACACCAAGCATGATCTGTGCGGCTTCTGTCAGGGGAACTCGATTTCTGCGAGGCTGAC 1433
 QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
 Db 1434 AGTGGAGGGCTCTGGTCACTTCGAGAACAAATATCTGGTGGCTGATAGGGATACAGC 1493
 QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1494 TGGGGTCTGGTGTGCCAAAGCTTACAGACGAGGAGTACGGGAATGTGATGGTATTC 1553
 QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
 Db 1554 ACGGACTGGATTATTCACAAATGAGGCGACGCGC 1589

```
RESULT 5
BC051839
LOCUS
DEFINITION
  Homo sapiens transmembrane protease, serine 2, mRNA (cDNA clone
  MGC:60144 IMAGE:6199625), complete cds.
ACCESSION
  BC051839
VERSION
  BC051839.1 GI:30353994
KEYWORDS
  MGC.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 3195)
  Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,
  Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
  Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
  Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
  Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
  Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
  Schmitt, T.E., Brownstein, M.J., Usdin, T.B., Tothiyuki, S.,
  Arancini, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
  Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J.,
  McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
  Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Fulyk, S.W.,
  Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
  Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
  Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
  Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
  Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
  Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,
  Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
  Generation and initial analysis of more than 15,000 full-length
  human and mouse cDNA sequences
  Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  22388257
  12477932
REFERENCE
  2 (bases 1 to 3195)
  Strausberg, R.
  Direct Submission
  Submitted (01-MAY-2003) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
  NIH-MGC Project URL: http://mgc.nci.nih.gov
  Contact: MGC help desk
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Dr. James R. Lupski
  CDNA Library Preparation: Life Technologies, Inc.
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Sequencing Group at the Stanford Human Genome
  Center, Stanford University School of Medicine, Stanford, CA 94305
  Web site: http://www-shgc.stanford.edu
  Contact: (Dickson, Mark) mcd@paxil.stanford.edu
  Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
  R. M.
  Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
  Series: IRAK plate: 110 Row: 0 Column: 3
  This clone was selected for full length sequencing because it
  passed the following selection criteria: matched mRNA gi: 14602458.
  Location/Qualifiers
    1..3195
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="MGC:60144 IMAGE:6199625"
      /tissue_type="Peripheral Nervous System, sympathetic
      trunk"
      /clone_lib="Lupski_sympathetic_trunk"
      /lab_host="DH103"
      /notes="Vector: pCMV-SPORT6"
    1..3195
FEATURES
  source
    gene
```


QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
Db 646 GGG 705
QY 201 ValAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 706 GTGGATGACAGCGGATCCACAGCTTTATGAACCTGAACACAGTGGCGGCAATGCGAT 765
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 766 ATCTATAAAAACTGTATCCACAGTATCCCTGTTCTTCAAAAGCAGTGGTTCTTACGC 825
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 826 TGTATAGCTCGGG 885
QY 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 886 AGCGCGCTCCGG 945
QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 946 TCGGAGGCTTCCATCATCACCCCGAGTGGATCGTGACAGCGGCCCATCGTGGGGA 1005
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1006 CCTCTTAACAATCCATGCGATTTGGAGCGCATTTGGGGGATTTTGACACATCTTT 1065
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1066 TTCTATGAGCGGGATACCAAGTGAAGAAGTATTTCTCATCAATATGATCTCAAG 1125
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1126 ACCAAGAACAATGACATTCGGCTGATGAAGCTGCAGAGCGCTCTGACTTTCAACG 1185
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1186 GTGAACACCATGTTCTGCTGCCAACCCAGCGCATGATGCTGCAGCGGACAGACCT 1245
QY 381 IleSerGlyTrpGlyAlaThrGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1246 ATTTCCGGTGG 1305
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1306 AAGGTGCTTCTCATTGACACACAGAGATGCAACAGCAGATATGCTATGACAACTG 1365
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1366 ACACAGCGCATGATCTGTCGGGGCTTCTGCGGGGGAACGTCGATCTTTCAGCGGTG 1425
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1426 AGTGAGGGGCTCTGGTCACTTCGAAGAACAATATCTGGTGGCTGATAGGGGATACA 1485
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1486 TGGGGTCTTGGCTGTGCCAAAGCTTACAGACAGAGGTGTCGGGAATGTGATGATTC 1545
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1546 ACGGACTGATTTATCGACAATAGGGGCGAGCGGC 1581

RESULT 6
AF123453
LOCUS Homo sapiens transmembrane serine protease 2 (TMPRSS2) mRNA, complete cds.
DEFINITION
ACCESSION AF123453
VERSION AF123453.1 GI:5001723
KEYWORDS
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 1479)
Teng,D.H., Chen,Y., Lian,L., Ha,P.C., Tavtigian,S.V. and Wong,A.K.
TITLE Mutation analyses of 268 candidate genes in human tumor cell lines
JOURNAL Genomics 74 (3), 352-364 (2001)
MEDLINE 21309069
PUBMED 11414763
REFERENCE 2 (bases 1 to 1479)
Teng,D.H.F., Chen,Y., Lian,L., Ha,P.C., Tavtigian,S.V. and Wong,A.K.C.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1999) Cancer Division, Myriad Genetics, Inc., 390 Wakara Way, Salt Lake City, UT 84108, USA
FEATURES
Location/Qualifiers
1..1479
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22"
/gene="TMPRSS2"
1..1479
/genes="TMPRSS2"
1..1479
/genes="TMPRSS2"
/codon_start=1
/product="transmembrane serine protease 2"
/protein_id="AAD37117.1"
/db_xref="GI:5001724"
translation="NALNSGSPALGPYYENHGYOPENYPYPAQTVPVTVYVHPAQY
YSPYPOVAPRVLTOASNPVCTOPKPSGTCTKTKKALCITLTLTGTLVGAALAA
GLLWPMGKSCNSGIECDSSGTCINPSNKGDSVGHSCGGEDENRCVRLYGNFLLQW
YSSQRKSHFVCOODWNNYGRACRDMYKKNFYSSQIGVDDSGTSFPMKNTSAGN
VDYIKLHSDACSKAVSLRDIACAGVNLNSRSRISVGGSSALPGAWGQYVLEHVQ
NVHVCSSGSIITPEWIVTAHCVEKPLNPNHWTAFAILRQSFYFGAGYQVEKVISH
PNYDSKNDIALMKLQKPLFNDLPKVPCLPNPGMILQPOLCWIISGWAETEKGK
TSEVLNAKVLIIETORNSRYVDNLITPAMICAGFLQGNVDSQCGDSGGPLVTSKN
NIWLIGDTSWSSGCAKVRPGVYGNVWFTDIWRQMRADG"

Alignment Scores:
Pred. No.: 5,31e-197 Length: 1479
Score: 2714.00 Matches: 491
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.80% Mismatches: 0
Query Match: 99.89% Indels: 0
DB: 9 Gaps: 0

US-09-615-285b-2 (1-492) x AF123453 (1-1479)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAACCTCAGGTGACCCAGCTATTGGACCTTACTATGAAACCATGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCGACAGCCACCTGCTGCCCTCCCTGCTACGAGGTGCAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACACCGCTCCCGGCTGCCCTGCGCCAGTACGCCCGAGGCTCTGACGCG 180
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCGCTCTCTGCACGACGCCAATCCCATCCCGGACAGTGTGCACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
Db 241 ACTAAGAAGACCTGTGCATCCTTGACCTGGACCTTCTCTGTTGGAGAGTGGCGCTG 300
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120

301 GCCCTGGCCTACTCTGGAAGTTCATGGGAGCAAGTGCTCCAACTCTGGGATAGAGTGC 360
121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
361 GACTCTCAGGTACTGATCAACCCCTCTAATCTGGTGTGATGCGGTGCACACTGCCCC 420
141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPhelLeuGlnVal 160
421 GCGGGGAGGACGAGATCGGTGTGTGCTCTACGACCAAACTTCATCTTCAGAT 480
161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
481 TACTCATCTCAGAGAGTCTGACACCTGTGTGCCAAGCAGCTGGAACGAGAACTAC 540
181 GlyArgAlaAlaCysArgAspMetGlyTyrIlysAsnAsnPhetYrSerSerGlnGlyIle 200
541 GCGGGGCGGCTCAGGAGCATGGCTATAAGAAATAATTTTACTCTAGCCCAAGGATA 600
201 ValAspAspSerGlySerThrSerPheMetIysLeuAsnThrSerAlaGlyAsnValAsp 220
601 GTGGATGACAGCGGATCACCAGCTTTATGAATGACACAGTGGCGGCAATGTCAT 660
221 IleTyrIlysLeuTyrHisSerAspAlaCysSerSerIysAlaValValSerLeuArg 240
661 ATCTATAAAACTGTACCAGATGATGCTGTTCTTCAAAAGCAGTGGTTCCTTACGC 720
241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
721 TGTATAGCTCGGGGTCACTTGAACCTCAAGCCGACAGAGTGGTGGGGGGTGGAG 780
261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
781 AGCGGCTCCCGGGGGCTGCGCTGGCAGTGCAGCTGCAGCTCCAGACGTCCTCAGTG 840
281 CysGlySerIleIleThrProGluTyrIleValThrAlaAlaHisCysValGlyLys 300
841 TCGGAGGCTCATCATCACCCCGAGTGGATCTGTGACAGCGCCCACTGCGTGGAAAA 900
301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
901 CCTCTTAAATPCCATGGCATTTGAGCGGCAATTTGCGGGATTTTGAGACATCTTTCTG 960
321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerIys 340
961 TTCTATGAGCGGATACCAAGTAGAAAAAGTATTTCTCATCCAAATTTAGCTCCAA 1020
341 ThrIysAsnAsnAspIleAlaLeuMetIysLeuGlnLysProLeuThrPheAsnAspLeu 360
1021 ACCAAGAACATGACATTCGCTGATGAAGCTGCAGAGCTCTGACTTTCAGACCTA 1080
361 ValIysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
1081 GTGAAACAGTGTGTCTGCCCAACAGCATGATGCTGCAGCCAGAACAGCTCTGCTGG 1140
381 IleSerGlyTyrGlyAlaThrGluLysGlyIysThrSerGluValLeuAsnAlaAla 400
1141 ATTTCCGGGTGGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTCTGAAAGCTGCC 1200
401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
1201 AAGGTGCTTCTCATTTGAGACACAGAGATGCAACAGCATATGTCTATGACAACTGATC 1260
421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
1261 ACACCAAGCCATGATCTGTCCCGGCTTCTGACAGGAGAAAGCTGATCTTGCAGGGTAC 1320
441 SerGlyGlyProLeuValThrSerIysAsnAsnIleTyrTrpLeuIleGlyAspThrSer 460
1321 AGTGGAGGGCCCTCTGTGTCTCTCGAAGAACATATCTGTTGGTGTATAGGGGATACAGC 1380
461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
1381 TGGGGTCTGCTGCCAAAGCTTACAGCAGGAGTGTACGGGAATGTGATGTTATTC 1440

481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
1441 ACGGACTGGATTTATCGACAAATGAGGACACGGC 1476

RESULT 7
AF329454 3226 bp mRNA linear PRI 16-MAY-2001
LOCUS Homo sapiens epitheliasin (TMPRSS2) mRNA, complete cds.
DEFINITION AF329454
ACCESSION AF329454.1 GI:14091027
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3226)
AUTHORS Jacquinet,E., Rao,N.V., Rao,G.V., Zhengming,W., Albertine,K.H. and Hoidal,J.R.
TITLE Cloning and characterization of the cDNA and gene for human epitheliasin
JOURNAL Eur. J. Biochem. 268 (9), 2687-2699 (2001)
MEDLINE 21232025
PUBMED 11322890
REFERENCE 2 (bases 1 to 3226)
AUTHORS Jacquinet,E., Rao,N.V., Rao,G.V., Wang,Z., Albertine,K.H. and Hoidal,J.R.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-2000) Pulmonary Medicine, University of Utah, SON. Medical Dr., Salt Lake City, UT 84132, USA
FEATURES
source
1..3226
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
1..3226
/gene="TMPRSS2"
128..1606
/gene="TMPRSS2"
/note="proteinase; PRSS10"
/codon_start=1
/product="epitheliasin"
/protein_id="AKS3559.1"
/db_xref="GI:14091028"
/translation="MALNSGPPAIGPYENHYQFENPYPAQPTVFTVYEVHPAQY
YSPVPQAPRVLTQASNPVCTQKSPSTVCTSKTKKALCITLTGLTFLVGAALAA
GLLWFMGSKSGTICDSGTINPNCWCDGSHCPGEGEDENCRVLYGNFLIQM
YSSQRKHPVCQDDWENYGRACKDMYKKNFYSSQIVDDSGTSPMKLNTSAGN
VDLYKLVHSDACSSKAVVSLACVGNLNSRSRIVGGSALPGAWPQVSLHVQ
NVHVCSSITPEWIVTAACVKEPLNPHHTAFAGILRQSFYMGAGYQVEKVIH
PNYDSTKNDIALMKLQKPLTFNDLVKFCVCLPNPQMWLOPQLQWISGAGTEKKG
TSEVLNAKVLIIETORCNSRYVDNLIITPAMICAGFLQGNVDSQGDSSGGPLVTSKN
NIWLIGDTSWGSCKAYRPGVYGVNMVFTDWIYRQMRADG"

Alignment Scores: 2,28e-196 Length: 3226
Pred. No.: 2711.00 Matches: 490
Score: 100.00% Conservative: 2
Percent Similarity: 99.59% Mismatches: 0
Best Local Similarity: 99.78% Indels: 0
Query Match: 9
DB: 0
US-09-615-285B-2 (1-492) x AF329454 (1-3226)

Qy 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 128 ATGGCTTGAACCTCAGGTCACCCAGCTATTGGACCTTACTATGAAACCAACGATGATC 187
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40

Db 188 CAACGGAAAAACCCCTATCCCGCACAGCCCACTGTGGTCCCCACTGTCTACGAGGTGCAT 247
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 248 CCGACTCAGTACTACCGTCCCGCGTCCCGCGTCCCGCGTCCCGCGTCCCGCGTCCCGCGT 307
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 308 TCCAAACCCGCTGTGTGTCAGCAGCCCAAAATCCCACTCCCGGACAGTGTGCATCCTCAAAG 367
Qy 81 ThrLysLysAlaLeuValCysThrLeuThrLeuThrLeuThrLeuThrLeuThrLeuThrLeu 100
Db 368 ACTAAGAAAGCAGTGTGCATCCTGACCTGACCTGCGGACCTTCTCTGCGGAGCTGCCTG 427
Qy 101 AlaAlaGlyLeuLeuThrLysPheMetGlySerLysCysSerAsnSerGlyLeuGluCys 120
Db 428 GCCGCTGCCCTACTCTGGAAGTTCATGGGACAGCAAGTCTCCAACTCTGGGATAGATGC 487
Qy 121 AspSerSerGlyThrCysLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 488 GACTCCTCAGTACTGTGCATCAACCCCTCTAACTGGTGTGATGGCGTGTGCATCTGCCCC 547
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 548 GCGGGGAGGAGCAGCAATCGTGTGCTGCTTACGACCAAACTTCATCCTTCAGATG 607
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 608 TACTCATCTCAGAGGAAGTCTCGCACCTGTGCGCAAGCAGCTGGAACGAGAACTAC 667
Qy 181 GlyArgAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyLe 200
Db 668 GCGCGGGCGGCTCGAGGACATGGGCTATGAAGATATATTTTACTCTAGCCAGGATA 727
Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 728 GTGATGACAGCGGATCCACAGCTTATGAAGCTGAACACAGTGGCGGCAATGTGCGAT 787
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
Db 788 ATCTATAAAACATGTATCACAGTATGCTTCTTCCAAAGCAGTGGTTCCTTACGC 847
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
Db 848 TGTATAGCTCGGGGTCACTTGAATCAAGCCGACAGCAGGATTTGGGCGGTGAG 907
Qy 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 908 AGCGCGCTCCCGGGGCGCTGCGCTGGCAGGTACAGCTGCACGTCAGAACTCCACGTG 967
Qy 281 CysGlyGlySerIleIleThrProGlnTrpIleValThrAlaAlaHisCysValGluLys 300
Db 968 TCGGAGAGCTCCATCATCACCCCGAGTGGATCGTGACAGCCGCCCTGCGTGGAAAAA 1027
Qy 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1028 CCTCTTAACAATCCATGGCATTTGGACGGCAATTTGGGGGATTTTGAGACAATCTTTTCATG 1087
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1088 TTCTATGAGCGCGGATACCAAGTAGAAAAGTATTTCTCATCCAAATATGATCTCCAG 1147
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1148 ACCAAGACAATGACATTGCGCTGATGAAGCTGCAGAGAGCTCTGACCTTCAAGAGCCTA 1207
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1208 GTGAACACCATGTGTCTGCCCAACCCAGCATGATGCTGCAGCAGAAACACTCTGCTGG 1267
Qy 381 IleSerGlyTrpGlyValaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1268 ATTTTCGGGTGGGGGGCCACCGAGGAGAAAGGAGAGACTCTCAGAGTGTCTGAACGCTGCC 1327

Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1328 AAGGTGCTTCTCATTTAGACACAGAGATGCCAACAGACATATGTCTATGACACCTGATC 1387
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1388 ACACGAGCCATATCTGTGCGGCTCTCTGCGAGGGAAGCTGATTTCTTGCAGGGTGAC 1447
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1448 AGTGGAGGSCCTCTGGTCACTTCGAAGAACAATATCTGGTGGCTGATAGGGGATACAAGC 1507
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1508 TGGGGTTCCTGGCTGTGCCAAAGCTTACAGACAGAGGTACGGATGTGTGATTTTC 1567
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1568 ACGGACTGGATTTATGACAAATGAGGCGAGACGGC 1603

RESULT 8

AX201158 1476 bp DNA linear PAT 29-AUG-2001
LOCUS Sequence 788 from Patent WO0151633.
DEFINITION AX201158
ACCESSION AX201158
VERSION AX201158.1 GI:15390913
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Reed, S.G., Kalos, M.D., Fanger, G.R., Day, C.H., Retter, M.W.,
Stolk, J.A., Skeiky, Y.A., Wang, A. and Meagher, M.J.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0151633-A 788 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source
1. 1476
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 1,27e-196 Length: 1476
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 6 Gaps: 0
US-09-615-285B-2 (1-492) x AX201158 (1-1476)
Qy 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTGAACCTCAGGTCACCCAGCTATTGGACCTTACTATGAAAACCATGGATAC 60
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 61 CAACCGGAAAAACCCCTATCCCGCACAGCCCACTGTGGTCCCCACTCTCTACGAGGTGCAT 120
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCGTCCCCCGTGGCCAGTACGCCCGGAGGTCTCTGACGCGGCT 180
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCGCTGTCTGTCAAGCAGCCCAAAATCCCACTCCCGGACAGTGTGCACTCAAG 240

QY 81 ThrIysAlaLeuCysIleThrLeuLeuGlyThrPheLeuValGlyAlaLeu 100
Db 241 ACTAAGAAACATGTCACCTTACCTGGGACCTTCTCTGGGAGCTGGCG 300
QY 101 AlaAlaGlyLeuLeuTrpIysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCCGCTGGCCTACTCTGGAAAGTTTCATGGGCAGCAAGTCTCCAACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGTACCTGTCATCAACCTCTTAACCTGGTGATGGCTGCACACTGCCCC 420
QY 141 GlyGlyGluAspGluLeuAsnArgCysValArgLeuTrpGlyProAsnPheIleLeuGlnVal 160
Db 421 GGGGGGAGACGAGATCGGTGTGTGGCTCTACGGATCAAACTTCATCTTCAGGTG 480
QY 161 TyrSerSerGlnArgIysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTrp 180
Db 481 TACTCATCTCAGAGGAGTCTGGCACCCTGTGTGCAAGACACTGCAACGAGAACTAC 540
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerGlnGlyIle 200
Db 541 GGGGGGCGCCTGACGGGACATGGGCTATAGAAATAATTTTACTCTAGCCAGGAATA 600
QY 201 ValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACAGCGGATCCACAGCTTATGAACACTGAACACAGTAGTCGGCAATGTCGAT 660
QY 221 IleTyrIysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAACCTGTACCAAGTATGCTGTCTTCTCAAGCATGGTTCCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerArgGlnSerArgIleValGlyGlu 260
Db 721 TGTATAGCCTGCGGGTCAACTTGAACCTCAAGCGCGCAGAGAGGATGTGGCGCGCAG 780
QY 261 SerAlaLeuProGlyAlaTrpTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 781 AGCGCGCTCCCGGGGCTGGCCTGGCAGGTGAGCTGCACGTCCAGAACGTCACGTG 840
QY 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 841 TGGGAGGCTCCATCATCACCCCGAGTGATGCTGACAGCGCCCACTCGGTGGAAAA 900
QY 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACAATCCATGGCATTTGACGCGCATTTTGGGGGATTTTGACAACTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTTCTATGGAGCCGATACCAAGTAGAAGAGTGTCTCATCCAAATTATGACTCCAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACAATGACATTTGGCTGATGAAGCTGCAGAAAGCTCTGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLysCysTrp 380
Db 1081 GTGAACACGAGTGTCTGTGCCAACCCAGGATGATGCTGCAGCAGAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGTGGGGGGCCACCGAGAGAGAGGAGAACCTCAGAAAGTGTGAAACGCTGCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTrpValTyrAspAsnLeuIle 420
Db 1201 AAGGTGCTTCTCATTTGAGACACAGATGCAACAGCAGATGATGCTATGACAACTTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACAGCCATGATCTGTCCGGCTTCTGTGAGGGGAAACGTCGATTTCTGCCAGGGTGC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460

Db 1321 AGTGGAGGCTCTGGTCACTTCGAAGAACAAATATCTGGTGGCTGATAGGGATACAAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGTTCTGGCTGTGCAAGAGCTTACAGACAGAGTGTACGGGAATGTGATGTATTC 1440
QY 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1441 ACGACTGGATTTATCGACAAATCGAGGCGACAGCGC 1476
RESULT 9
AX267957 1476 bp DNA linear PAT 26-OCT-2001
LOCUS Sequence 931 from Patent WO0173032.
DEFINITION AX267957
ACCESSION AX267957
VERSION AX267957.1 GI:16516542
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Xu, J., Dallon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Kalos, M.D., Fanger, G.R., Retter, M.W., Stolk, J.A., Day, C.H.,
Vedvick, T.S., Carter, D., Li, S.X., Wang, A., Skeiky, Y.A., Hepler, W.T.
and Henderson, R.A.
TITLE Compositions and Methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0173032-A 931 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..1476
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores: 1.27e-196 Length: 1476
Pred. No.: 491
Score: 2709.00 Matches:
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: Gaps: 0
US-09-615-285B-2 (1-492) x AX267957 (1-1476)
QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAACCTCAGGGTCAACCACTATGGACCTTACTATGAACCACTGGATAC 60
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCCGCAGCCCACTGTGTCCTCCCACTGTCTACGAGGTGAT 120
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCGCCGTCAGTACGCCCGGAGGCTCTGACGAGGCT 180
QY 61 SerAsnProValValCysThrGlnProLysProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCGCTGTGACGAGCCCAATCCCATCCCGACAGTGTGACCTCAAG 240
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaLeu 100
Db 241 ACTAAGAAGCACTGTGCATCACCTTGACCTGGGACCTTCTCTGGGAGCTGCGCTG 300
QY 101 AlaAlaGlyLeuLeuTrpIysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCGCTGGCTACTCTGGAAGTTCATGGCAGCAAGTGTCCCACTCTGGGATAGAGTGC 360
QY 121 AspSerSerGlyThrCysIleLeuAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140

Db 361 GACTCCTCAGTACCTGATCAACCCCTTAACCTGGTGTGATGGCGTGTACACTGCCCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProbenPheLeuGlnVal 160
Db 421 GCGGGGAGGACGAGATCGGTGTGGCTCTACGATCAAACTTCTCTCAGGTG 480
Qy 161 TyrSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGAAAGTCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGACTAC 540
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrIysAsnAsnPheTyrSerGlnGlyLe 200
Db 541 GGGGGGGCGCTCGAGGACATGGCTATAAGAAATAATTTTACTCTAGCCAGGAATA 600
Qy 201 ValAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 601 GTGGATGACACGGATCCACAGCTTTATGAACCTGACACAGTGGCGGCAATGTGAT 560
Qy 221 IleTyrIysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 661 ATCTATAAAACCTGTACACAGTATGCTCTTCTTCAAAAGCAGTGGTCTTTTACGC 720
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyLe 260
Db 721 TGTATAGCTCGGGGTCACCTGAACTCAAGCGCGCAGAGAGAGATTTGGCGCGAG 780
Qy 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnValHisVal 280
Db 781 AGCGCGTCCGGGGCTGGCCCTGGCAGCTCAGCTGCAGCTCCAGACGCTCCACGTG 840
Qy 281 CysGlyGlySerIleLeuThrProGluTyrIleValThrAlaAlaHisCysValGluLys 300
Db 841 TCGGAGGTCCATCATACCCCGAGTGTATCGTACAGCCGCCACCTCGCTGGAAAGA 900
Qy 301 ProLeuAsnAsnProTyrHisTyrThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 901 CCTCTTAACATCCATGCGATTCGACGGCATTTGGCGGGATTTTGAGACAATCTTTCATG 960
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 961 TTCTATGAGCGCGGATCCAAAGTAGAAAGTAGATTTCTATCCAAATATGACTCCAA 1020
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1021 ACCAAGAACATGACATTTGGCTGATGAAGCTGCAGAGCCCTCTGACTTTCAACGACCTA 1080
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
Db 1081 GTGAACCATGTGTCTGCCCAACCCAGGATGATGCTGCAGCAGACAGCTCTGCTGG 1140
Qy 381 IleSerGlyTyrGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1141 ATTTCCGGTGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTGTGAAAGCTGCC 1200
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerAtgTyrValTyrAsnLeuIle 420
Db 1201 AAGTGCTTCTATTGAGACACAGATGATGATGATGATGATGATGATGATGATGATGATG 1260
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1261 ACACGACCATATCTGTGCGGCTTCTCTGAGGGAAAGCTGATTTTGCCAGGGTGCAC 1320
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
Db 1321 AGTGGAGGGCTCTGGTCACTTCGAAGAACAATATCTGGTGGCTGATAGGGGATACAAAG 1380
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1381 TGGGGTCTGGCTGTGCCAAAGCTTACACAGCAGGAGTGTACGGGAATGTGTGATTC 1440
Qy 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492

Db 1441 ACGGACTGGATTATCGCAAAATGAGGGCAGACGGC 1476
RESULT 10
AX201157
LOCUS AX201157
DEFINITION Sequence 787 from Patent WO0151633.
ACCESSION AX201157
VERSION AX201157.1 GI:15390912
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Reed, S.G., Kalos, M.D., Fanger, G.R., Day, C.H., Retter, M.W.,
Scolk, J.A., Skeiky, I.A., Wang, A. and Meagher, M.J.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0151633-A 787 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..1479
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 1,276-196 Length: 1479
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservatives: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
Gaps: 0
US-09-615-285B-2 (1-492) x AX201157 (1-1479)
Qy 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 1 ATGGCTTTGAACTCAGGGTCACACAGCTATTGGACCTTACTATGAAACCATGGATAC 60
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 61 CAACCGGAAACCCCTATCCGCACAGCCACCTGTGTGTCCTCCCTCTACGAGTGAT 120
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 121 CCGGCTCAGTACTACCCGTCCTCCCTGAGTACGAGGAGGCTCTGACGAGGCT 180
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 181 TCCAAACCCGCTGTCTGCACGACGACCAATCCCATCCGAGACAGTGTGCACCTCAAAG 240
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
Db 241 ACTAAGAAAGCACTGTGCATCACCTTGACCTTGGGAGCTTCTCTGTTGGAGCTGCGCTG 300
Qy 101 AlaAlaGlyLeuLeuTyrPhePheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 301 GCGCTGGGCTACTCTGAAAGTTATCGGCAGCAAGTCTCCAACTCTGGATAGAGTGC 360
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
Db 361 GACTCTCAGGTACTGTGATCAACCCCTCTAAGTGTGTGATGGGTGTACACTGCCCC 420
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 421 GCGGGGAGGACGAGATCGGTGTGTTGCTCTACGATCAAACTTCTCTCAGGTG 480
Qy 161 TyrSerSerGluArgLysSerTyrHisProValCysGlnAspTyrAsnGluAsnTyr 180
Db 481 TACTCATCTCAGAGAAAGTCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGACTAC 540

QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
DB 661 ATCTATAAAAACTGACACAGTATGCTGTTCTTCAAAAGCAGTGGTTCTTTACGC 720
QY 241 CysIleAlaCysGlyValAsnLeuSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 721 TGTATAGCTGCGGGGTCACTTGACTCAGCCCGCAGACAGAGTGTGGCGCGAG 780
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280
DB 781 AGCGCGCTCCCGGGGCGCTGGCCCTGGCAGGTGACCTGCACGTCCAGAGTCCACGTG 840
QY 281 CysGlyGlySerIleLeuThrProGlnThrPheValThrAlaAlaHisCysValGluLys 300
DB 841 TCGGAGGCTCCATCATCACCCCGAGTGGATCGACAGCGCCCATCTGGTGGAAAA 900
QY 301 ProLeuAsnAsnProTyrHisThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
DB 901 CCTCTTAAATCCATGTCAGTGGACGCAATTTGGCGGATTTTGAGACAATCTTTTCATG 960
QY 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAsnSerLys 340
DB 961 TTCTATGAGGCGGATACCAATAGAAAAAGTGAATTTCTCATCCAAATATGACTCCCAAG 1020
QY 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
DB 1021 ACCAAGACAATGACATTTGGCTGATGAAGCTGCAGAGCCCTTCGACTTTCAACGACCTA 1080
QY 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGlnLeuCysTyr 380
DB 1081 GTGAACACCATGTGTCTGCCCAACCCAGCATGATGCTGCAGCCAGAACAGCTCTGCTGG 1140
QY 381 IleSerGlyTyrGlyValThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
DB 1141 ATTTCCGGTGGGGGCCACCGAGAGAAAGGAGAGCCTCAGAGTGTCTGACCGTCC 1200
QY 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAsnLeuIle 420
DB 1201 AAGGTGCTTCTCATTCAGACACAGAGATGCAACAGCAGATATGTCTATGACCAACCTGATC 1260
QY 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
DB 1261 ACACAGCCCATGATCTGTCCCGCTTCTCGAGGGGAAGCTCGATTTCTGCCAGGGTGAC 1320
QY 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
DB 1321 AGTGGAGGCGCTCTGCTCACTTCGAGAACATAATCTGGTGGCTGATAGGGGATACAAGC 1380
QY 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
DB 1381 TGGGGTCTGGCTGTGCCAAAGCTTACAGACAGAGGTATACGGGAATGTGTATTC 1440
QY 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
DB 1441 ACGGACTGATTTATCGACAATGAGGCGAGCGGC 1476

RESULT 12
AX201156 LOCUS 3245 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 786 from Patent WO0151633.
ACCESSION AX201156
VERSION AX201156.1 GI:15390911
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.
TITLE Compositions and methods for the therapy and diagnosis of prostate cancer

JOURNAL Patent: WO 0151633-A 786 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
source 1. 3245
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 3,25e-196 Length: 3245
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x AX201156 (1-3245)

QY 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
DB 145 ATGGCTTTGAACTCAGGTCACCCAGCTATTGGACCTTACTATCAAAACCATGGATAC 204
QY 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
DB 205 CAACCGGAAACCCCTATCCGCACAGCCACTGTGTCCTCCCACTGTCTACGAGGTGCAT 264
QY 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
DB 265 CCGGCTCAGTACTACCCGTCCTCCGTCGCCCAGTACGCCCGGAGGTCTCTGACGAGCT 324
QY 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
DB 325 TCCAACCCCTGCTCTGCACGACGAGCCAAATCCCGACAGTGTGCACCTCAAG 384
QY 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
DB 385 ACTAAGAAAGACATGTGCATCCTTGACCTCGGGACCTTCTCTCGTGGAGCTGCGTG 444
QY 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
DB 445 GCGCTGCGCTACTCTGGAAGTTTCATGGGCAGCAAGTCTCCAACTCTGGGATAGAGTGC 504
QY 121 AspSerSerGlyThrCysIleAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
DB 505 GACTCTTCAGGTACCTGCATCAACCCCTCTAATCTGGTGTGATGGCTGTCTCACACTGCC 564
QY 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
DB 565 GCGGGGAGGACGAGATCGTGTGTTGCTCTTACGGATCAAACTTCACTCTAGCGGTG 624
QY 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
DB 625 TACTCATCTCAGAGGAAGTCTTGGCACCCCTGTGTGCCAAGCAGCTGGAACGAGAACTAC 684
QY 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
DB 685 GGGCGGGCGCTTCAGGACATGGGTATAGAAATAATTTTACTCTAGCCCAAGGAATA 744
QY 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
DB 745 GTGGATCAGACGCGATCCACAGCTTTATGAAATGAACTGAACAACAAGTCCCGCAATGTGCAT 804
QY 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValSerLeuArg 240
DB 805 ATCTATATAAAACCTGTACCACTGATGCTCTTCTTCAAAAGCAGTGGTTCCTTTACGC 864
QY 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
DB 865 TGTATAGCTCGCGGGTCAACTTGAATCAAGCCCGCAGAGGATTTGTGGCGCGGAG 924
QY 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnAsnValHisVal 280

Db 925 AGCGCGCTCCGGGGGCTGGCCCTGGCAGGTGAGCTGACGTCCAGAACGTCCACGTG 984
Qy 281 CysGlySerIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 985 TGGGAGGCTCCATCATCACCCCGAGTGGATCTGACAGCGCCCGCTCGGTGGAAAA 1044
Qy 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1045 CCTCTTAACAATCCATGGCATTCGACGGCATTTGGCGGGATTTTGAGACAATCTTTTCATG 1104
Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
Db 1105 TTCTATGGAGCCGGATACCAAGTAGAAAAAGTGAATTTCTCATCCAAATTTATGCTCCAG 1164
Qy 341 ThrLysAsnAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
Db 1165 ACCAAGAACCAATGACATTCGGCTGATGAAGCTGCAGAACCTCTGACTTTTCAACGACCTA 1224
Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTrp 380
Db 1225 GTGAACCAAGTGTCTGCCCCAACCCAGGATGATGCTGACGACAGAACAGTCTGCTGG 1284
Qy 381 IleSerGlyTrpGlyAlaThrGluGluLysGlyLysThrSerGluValLeuAsnAlaAla 400
Db 1285 ATTTCGGGTGGGGGCCACCGAGGAGAAAGGAGAACCTCAGAGTGTCTGAACGCTGCC 1344
Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeuIle 420
Db 1345 AAGTGCTTCTCATTTAGACACAGAGATGCAACAGCAGATATCTCTATGACACCTTGATC 1404
Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
Db 1405 ACACCACCATGATCTGCGCGCTTCTCGAGGGAACGTCTGATTTTCCAGGCTGAC 1464
Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTrpTrpLeuIleGlyAspThrSer 460
Db 1465 AGTGGAGCGCTCTGTGTCACCTTCGAAGACAATATCTGTGGTGTATGAGGGATACAGC 1524
Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
Db 1525 TGGGGTTCTGCTGTGCAAGCTTACAGACCGAGTGTACGGGATGTGATGTGATTC 1584
Qy 481 ThrAspTrpIleTyrArgGlnMetArgAlaAspGly 492
Db 1585 ACAGCTGGATTTATCGACAAATGAGGGCAGACGGC 1620

RESULT 13
AX267955 3245 bp DNA linear PAT 26-OCT-2001
LOCUS
DEFINITION Sequence 929 from Patent WO0173032.
ACCESSION AX267955
VERSION AX267955.1 GI:16516540
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
XU,J., DILLON,D.C., MITCHAM,J.L., HARLOCKER,S.L., JIANG,Y.,
KALOS,M.D., FANGER,G.R., RETTER,M.W., STOLK,J.A., DAY,C.H.,
VEDVICK,T.S., CARTER,D., LI,S.X., WANG,A., SKEIKY,Y.A., HEPLER,W.T.
and HENDERSON,R.A.
Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0173032-A 929 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
source 1..3245
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 3,25e-196 Length: 3245
Score: 2709.00 Matches: 491
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 99.71% Indels: 0
DB: 6 Gaps: 0
US-09-615-285b-2 (1-492) x AX267955 (1-3245)

Qy 1 MetAlaLeuAsnSerGlySerProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
Db 145 ATGGCTTTGAATCAGGGTCACACCAAGTATTGACCTTACTATGAAACACATGATAC 204
Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValProThrValTyrGluValHis 40
Db 205 CAACCGGAAAAACCCCTATCCCGCAGCCCACTGTGTGCCCTACTGTCTACGAGGTGAT 264
Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
Db 265 CGGGCTCAGTACTACCCGTGCCCGTCCCGCAGTACGCCCGAGGGTCTCTGACGAGCT 324
Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
Db 325 TCCAAACCCCGTCTGTGCACGACGCCCAATCCCCATCCCGGACAGTGTGCACCTCAAAG 384
Qy 81 ThrLysLysAlaLeuCysIleThrLeuThrLeuGlyThrPheLeuValGlyValAlaLeu 100
Db 385 ACTAAGAAGACATGTGCATCACTTGACCTGGGACCTTCTCTCGTGGAGCTGCGCTG 444
Qy 101 AlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
Db 445 GCGGTGCGCTTACTCTGGAAGTTTATGGCAGCAAGTGTCTCAACTCTTGGGATAGAGTGC 504
Qy 121 AspSerSerGlyThrCysIleAsnProSerAsnTrpCysAspGlyValSerHisCysPro 140
Db 505 GACTCTCAGGTACTCTGTCATCAACCCCTCTAACTGTGTGTATGCGGTGTACACTGCC 564
Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
Db 565 GCGGGGAGGACGAGATCGGTGTGTCTGCTTACGGATCAAACTTCTATCTTCAGGTG 624
Qy 161 TyrSerSerGlnArgLysSerTrpHisProValCysGlnAspAspTrpAsnGluAsnTyr 180
Db 625 TACTCATCTCAGAGGAAGTCTGTCACCTGTGCCAAGACGACTGGAACGAGNACTAC 684
Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
Db 685 GGGCGGGCGGCTGCAGGACATGGGCTATAAGAATAATTTTACTCTAGCCAAGGAATA 744
Qy 201 ValAspAspSerGlySerThrPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
Db 745 GTGGATGACAGCGGATCCACCACTTTATGAACCTGAAACACAGTCCCGCAATGTGAT 804
Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
Db 805 ATCTATAAATAACTGTACACAGTATGCTTCTTCAAAAGCAGTGGTTCCTTTCTAGCC 864
Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlu 260
Db 865 TGTATAGCTTGGCGGGTCAACTTGAACCTCAAGCCGCCAGACGAGATTTGTGGCGGCGAG 924
Qy 261 SerAlaLeuProGlyAlaTrpProTrpGlnValSerLeuHisValGlnAsnValHisVal 280
Db 925 AGCGGCTCCCGGGGCTGTGCCCTGGCAGGTGACCTGCACGTCCAGACGTCACACGTG 984
Qy 281 CysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLys 300
Db 985 TCGCGAGGCTCCATCATACCCCGAGTGGATCTGTGACAGCGCGCCCTGCTGGTGAATAA 1044
Qy 301 ProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMet 320
Db 1045 CCTCTTAACAATCCATGGCATTTGAGGCAATTTGAGACATCTTTTCATG 1104

Qy 301 ProLeuAsnProTyrHisThrAlaPheAlaGlyLeuArgGlnSerPheVal 320
 Db 1045 CCTTAAACATCAATGGCATTTGGCGGATTTGGGGATTTTGGAGCAATCTTTCATG 1104
 Qy 321 PheTyrGlyAlaGlyTyrGlnValGluLysValIleSerHisProAsnTyrAspSerLys 340
 Db 1105 TTTCTATGGAGCGGATACCAAGTACAAAGATGATTTCTCATCCAAATATGACTCCAG 1164
 Qy 341 ThrLysAsnAspIleAlaLeuMetLysLeuGlnLysProLeuThrPheAsnAspLeu 360
 Db 1165 ACCAAGAACATGACATTCGCTGATGAAGCTGCGAAGCCCTCAGACTTCAACGACCTA 1224
 Qy 361 ValLysProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTyr 380
 Db 1225 GTGAACACAGTGTCTGCCACCCAGGATGATGCTGCAGCCAGACAGCTCTGCTGG 1284
 Qy 381 IleSerGlyTyrGlyAlaThrGluGlnLysGlyLysThrSerGluValLeuAsnAla 400
 Db 1285 ATTTCGGGTGGGGGGCCACCGAGGAGAAAGGAGAGCTCAGAAAGTCTGAACGCTGCC 1344
 Qy 401 LysValLeuLeuIleGluThrGlnArgCysAsnSerArgTyrValTyrAspAsnLeu 420
 Db 1345 AAGGTGCTTCTCATTCAGACACAGAGATGCAACACAGATATGCTATGACAACTGATC 1404
 Qy 421 ThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAsp 440
 Db 1405 ACACCAAGCATGATCTGTCGGCTCTCTGCAGGAGAAAGCTGATCTTTCGCGGGTGAC 1464
 Qy 441 SerGlyGlyProLeuValThrSerLysAsnAsnIleTyrTyrLeuIleGlyAspThrSer 460
 Db 1465 AGTGGAGGGCTCTGCTCACTTCGAAGAACATATCTGGTGGCTGATAGGGGATACAGC 1524
 Qy 461 TrpGlySerGlyCysAlaLysAlaTyrArgProGlyValTyrGlyAsnValMetValPhe 480
 Db 1525 TGGGGTCTGGCTGTGCCAAGCTTACAGACAGAGGTGTACGGGAATGTGATGATTC 1584
 Qy 481 ThrAspTyrIleTyrArgGlnMetArgAlaAspGly 492
 Db 1585 ACGGACTGGATTATCGACAAATGAGGGCAGACGGC 1620

RESULT 15
 BD228306
 LOCUS
 DEFINITION
 Method of diagnosis, observation, staging, imaging and treatment of
 prostatic cancer.
 ACCESSION
 BD228306.1 GI:33038076
 VERSION
 JP 2002527758-A/8.
 KEYWORDS
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Suteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 Salceda, S., Recipon, H. and Caferkey, R.
 Method of diagnosis, observation, staging, imaging and treatment of
 prostatic cancer.
 JOURNAL
 Patent: JP 2002527758-A 8 27-AUG-2002;
 DIADEX INC
 COMMENT
 OS Homo sapiens (human)
 PN JP 2002527758-A/8
 PD 27-AUG-2002
 PF 19-OCT-1999 JP 2000576884
 PR 19-OCT-1998 US 60/104737
 PI SUSANA SALCEDA, HERVE RECIPON, ROBERT CAFFERKEY PC
 G01N33/574, A61K39/395, A61K39/395, A61K49/00, A61K51/00, A61P35/00, PC
 C07K16/32,
 C12N15/09, C12Q1/68, G01N33/577, A61K49/02, C12N15/00 CC Method
 of diagnosis, observation, staging, imaging and CC
 treatment of
 CC prostatic cancer
 FH Key Location/Qualifiers
 FT source I. .3443

FT
 FEATURES
 source
 Location/Qualifiers
 1. .3443
 /organism="Homo sapiens"
 /mol_type="Genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores:
 Pred. No.: 8,378-196 Length: 3443
 Score: 2704.00 Matches: 492
 Percent Similarity: 99.80% Conservative: 0
 Best Local Similarity: 99.80% Mismatches: 0
 Query Match: 99.52% Indels: 1
 DB: 6 Gaps: 0

US-09-615-285B-2 (1-492) x BD228306 (1-3443)

Qy 1 MetAlaLeuAsnSerGlySerProProAlaIleGlyProTyrTyrGluAsnHisGlyTyr 20
 Db 149 ATGGCTTTGAACCTCAGGGTCAACCAAGCTATTGGACCTTACTATGAAACCACTGGATAC 208
 Qy 21 GlnProGluAsnProTyrProAlaGlnProThrValValProThrValTyrGluValHis 40
 Db 209 CAACCGGAAACCCCTATCCCGCACAGCCACTGTGGTCCCCACTGTCTACGAGGTGCAT 268
 Qy 41 ProAlaGlnTyrTyrProSerProValProGlnTyrAlaProArgValLeuThrGlnAla 60
 Db 269 CGGGCTCAGTACTACCGTCCCGGTGCCCGAGTACGCCCGAGGGTCTTGACGAGGCT 328
 Qy 61 SerAsnProValValCysThrGlnProLysSerProSerGlyThrValCysThrSerLys 80
 Db 329 TCCAAACCCGCTGCTGACGACGACCAAAATCCCAATCCCGGAGAGTGTGCACCTCAAAG 388
 Qy 81 ThrLysLysAlaLeuCysAlaThrLeuThrLeuGlyThrPheLeuValGlyAlaAlaLeu 100
 Db 389 ACTAAGAAGACACTGTGCATCACCTTGACCTGGGACCTTCTCTGGGAGGTGCGCTG 448
 Qy 101 AlaAlaGlyLeuLeuTyrLysPheMetGlySerLysCysSerAsnSerGlyIleGluCys 120
 Db 449 GCGGTGGCTTACTCTGAAGTTTATGGGAGCAAGTGTCCCACTCTGGGATAGAGTGC 508
 Qy 121 AspSerSerGlyThrCysAlaAsnProSerAsnTyrCysAspGlyValSerHisCysPro 140
 Db 509 GACTCCTCAGGTACTGTGCATCAACCCCTTACTTGTGTGTATGGGTGTGCACACTGCC 568
 Qy 141 GlyGlyGluAspGluAsnArgCysValArgLeuTyrGlyProAsnPheIleLeuGlnVal 160
 Db 569 GCGGGGAGGACGAGATCGGTGTGCTTCCCTCTACGACCAAACTTCATCCITCAGGTG 628
 Qy 161 TyrSerSerGlnArgLysSerTyrHisProValCysGlnAspAspTyrAsnGluAsnTyr 180
 Db 629 TACTCATCTCAGAGGAAGTCTTGGCACCCCTGTGTGCAAGACGACTGGAACGAGACTAC 688
 Qy 181 GlyArgAlaAlaCysArgAspMetGlyTyrLysAsnAsnPheTyrSerSerGlnGlyIle 200
 Db 689 GCGCGGGGGCTCGAGGACATGGGTATTAAGATTAATTTTACTCTACCAAGGAATA 748
 Qy 201 ValAspAspSerGlySerThrSerPheMetLysLeuAsnThrSerAlaGlyAsnValAsp 220
 Db 749 GTGGATGACAGCGGATCCACCGCTTTATGAAACTGAACAAGTGCAGCAATGTTCAT 808
 Qy 221 IleTyrLysLysLeuTyrHisSerAspAlaCysSerSerLysAlaValValSerLeuArg 240
 Db 809 ATCTATAAAACATGTACCAAGTGTCTTCTTCAAAGACAGAGGTGTTCTTTTACGC 868
 Qy 241 CysIleAlaCysGlyValAsnLeuAsnSerSerArgGlnSerArgIleValGlyGlyGlu 260
 Db 869 TGTATAGCTCGGGGTCAACTTGAACCTCAAGCGCCGAGAGGATCGTGGCGGCGAG 928
 Qy 261 SerAlaLeuProGlyAlaTyrProTyrGlnValSerLeuHisValGlnValHisVal 280
 Db 929 AGCGCTCTCCGGGGGCTGCGCTTGGGAGGTGAGCTTCACAGACCTCCAGACCTCCACGT 988

QY 280 lCysGlyGlySerIleIleThrProGluTrpIleValThrAlaAlaHisCysValGluLy 300
Db |||||
989 GTGCGAGGCTCCATCATCACCCCGAGTGGATCGTGACGCGCCCACTGCTGGAAAA 1048
QY 300 sProLeuAsnAsnProTrpHisTrpThrAlaPheAlaGlyIleLeuArgGlnSerPheMe 320
Db |||||
1049 ACCTCTTAACAATCCATGSCATTGGACGGCAATTCGGGGATTTTGACACAATCTTTCAT 1108
QY 320 tPheTyrglyAlaGlyTyrglnValGluLyValIleSerHisProAsnTyraSpSerLy 340
Db |||||
1109 GTTCTATGGAGCGGATACCAAGTAGAAAAAGTAGATTCTCATCCAAATTATGACTCCAA 1168
QY 340 sThrLyAsnAsnAspIleAlaLeuMetLyLeuGlnLyProLeuThrPheAsnAspLe 360
Db |||||
1169 GACCAAGACATGACATTCGCTCATGAGCTGCAGAGCCTCTGACTTTCACAGACCT 1228
QY 360 uValLyProValCysLeuProAsnProGlyMetMetLeuGlnProGluGlnLeuCysTr 380
Db |||||
1229 AGTGAACCCAGTGTCTCTGCCAACCCAGGCATGATGCTGCAGCCAGAACAGCTCTGCTG 1288
QY 380 pIleSerGlyTrpGlyAlaThrGluGluLyGlyLyThrSerGluValLeuAsnAlaAl 400
Db |||||
1289 GATTTCCGGGTGGGGGCCACCGAGGAGAAAGGGAGACCTCAGAACTGCTGAACGCTGC 1348
QY 400 aLyValLeuLeuIleGluThrGlnArgCysAsnSerArgTyValTyraSpAsnLeuIl 420
Db |||||
1349 CAAGGTGCTTCTCATTGACACACAGAGATGCAACACAGATATGTCTATGACACCTCAT 1408
QY 420 eThrProAlaMetIleCysAlaGlyPheLeuGlnGlyAsnValAspSerCysGlnGlyAs 440
Db |||||
1409 CACACAGCCATGATCTGTGCCGGCTTCTGCAGGGGAACGTCGATTCTTGCCAGGGTGA 1468
QY 440 pSerGlyGlyProLeuValThrSerLyAsnAsnIleTrpTrpLeuIleGlyAspThrSe 460
Db |||||
1469 CAGTGGAGGGCCTCTGGTCACCTCGAGAGACATATCTGGTGGCTGATAGGGGATACAAG 1528
QY 460 rTrpGlySerGlyCysAlaLyAlaTyraArgProGlyValTyrglyAsnValMetValPh 480
Db |||||
1529 CTGGGGTCTGGCTGTGCCAAAGCTTACAGACCGAGGTACGGGAATGTGATGGTATT 1588
QY 480 eThrAspTrpIleTyraGlnMetArgAlaAspGly 492
Db |||||
1589 CACGACTGGATTATCGACAAATGAGGGCAGACGGC 1625

Search completed: June 1, 2004, 16:37:10
Job time : 5898 secs

; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 586
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-586

Query Match 100.0%; Score 2717; DB 15; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.1e-230;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGPPAIGYENHGYQENPYPAQPTVTVYEVHPAQYPSVPQYAPRVLTQA 60
DB 1 MALNSGPPAIGYENHGYQENPYPAQPTVTVYEVHPAQYPSVPQYAPRVLTQA 60

QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLWKFMSKCSNSGIEC 120
DB 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLWKFMSKCSNSGIEC 120

QY 121 DSSGTCINPNCWCDGVSHCPGGEDENRCVRLYGPNFILOYYSQKSWHPVCQDDWNEY 180
DB 121 DSSGTCINPNCWCDGVSHCPGGEDENRCVRLYGPNFILOYYSQKSWHPVCQDDWNEY 180

QY 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240

QY 241 CIACGVNLNRSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWITVAACHVEK 300
DB 241 CIACGVNLNRSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWITVAACHVEK 300

QY 301 PLNPNWHTAFAGILRQSFYAGYQVEKVIHSPNYDSKTNNDIALMKLOKPLTFNDL 360
DB 301 PLNPNWHTAFAGILRQSFYAGYQVEKVIHSPNYDSKTNNDIALMKLOKPLTFNDL 360

QY 361 VKPCLPNPQWMLQPEQLCWISGKGATEEKGKTSSEVLNAKVLLIETQCNRSRYVDNLI 420
DB 361 VKPCLPNPQWMLQPEQLCWISGKGATEEKGKTSSEVLNAKVLLIETQCNRSRYVDNLI 420

QY 421 TPAMICAGFLQGNVDSQCGSGGLPVTSKNNIWWLIGDTSWGSCKAKAYRPGYGNVMVF 480
DB 421 TPAMICAGFLQGNVDSQCGSGGLPVTSKNNIWWLIGDTSWGSCKAKAYRPGYGNVMVF 480

QY 481 TDWIYRQWRADG 492
DB 481 TDWIYRQWRADG 492

RESULT 2
US-10-205-823-415
; Sequence 415, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205.823

; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 415
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-415

Query Match 99.8%; Score 2711; DB 14; Length 492;
Best Local Similarity 99.6%; Pred. No. 1.7e-229;
Matches 490; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGPPAIGYENHGYQENPYPAQPTVTVYEVHPAQYPSVPQYAPRVLTQA 60
DB 1 MALNSGPPAIGYENHGYQENPYPAQPTVTVYEVHPAQYPSVPQYAPRVLTQA 60

QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLWKFMSKCSNSGIEC 120
DB 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLWKFMSKCSNSGIEC 120

QY 121 DSSGTCINPNCWCDGVSHCPGGEDENRCVRLYGPNFILOYYSQKSWHPVCQDDWNEY 180
DB 121 DSSGTCINPNCWCDGVSHCPGGEDENRCVRLYGPNFILOYYSQKSWHPVCQDDWNEY 180

QY 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240

QY 241 CIACGVNLNRSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWITVAACHVEK 300
DB 241 CIACGVNLNRSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWITVAACHVEK 300

QY 301 PLNPNWHTAFAGILRQSFYAGYQVEKVIHSPNYDSKTNNDIALMKLOKPLTFNDL 360
DB 301 PLNPNWHTAFAGILRQSFYAGYQVEKVIHSPNYDSKTNNDIALMKLOKPLTFNDL 360

QY 361 VKPCLPNPQWMLQPEQLCWISGKGATEEKGKTSSEVLNAKVLLIETQCNRSRYVDNLI 420
DB 361 VKPCLPNPQWMLQPEQLCWISGKGATEEKGKTSSEVLNAKVLLIETQCNRSRYVDNLI 420

QY 421 TPAMICAGFLQGNVDSQCGSGGLPVTSKNNIWWLIGDTSWGSCKAKAYRPGYGNVMVF 480
DB 421 TPAMICAGFLQGNVDSQCGSGGLPVTSKNNIWWLIGDTSWGSCKAKAYRPGYGNVMVF 480

QY 481 TDWIYRQWRADG 492
DB 481 TDWIYRQWRADG 492

RESULT 3
US-09-759-143-932
; Sequence 932, Application US/09759143
; Patent No. US200202248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.

APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C23
CURRENT APPLICATION NUMBER: US/09/759,143
CURRENT FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
TYPE: PRT
LENGTH: 492
ORGANISM: Homo sapiens
US-09-759-143-932

Query Match 99.7%; Score 2709; DB 9; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MALNSGSPALPGYENHGYQENPYPAQPTVTVTVYEVHPAQYYPSPVQYAPRVLTQA 60
Db 1 MALNSGSPALPGYENHGYQENPYPAQPTVTVTVYEVHPAQYYPSPVQYAPRVLTQA 60
Qy 61 SNPVVCTOPKSPSGTCTKTKALCITLTGTLVGAALAAAGLLWKFMSKCSNSGIEC 120
Db 61 SNPVVCTOPKSPSGTCTKTKALCITLTGTLVGAALAAAGLLWKFMSKCSNSGIEC 120
Qy 121 DSSGTCINPSNWCDCVSHCPGGEDENRCVRLYGNFILOVYSSORKSWHPVCQDDWNEY 180
Db 121 DSSGTCINPSNWCDCVSHCPGGEDENRCVRLYGNFILOVYSSORKSWHPVCQDDWNEY 180
Qy 181 GRAACRDMGYKNFYSSQGIIVDDSGTSFMKLNLSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGIIVDDSGTSFMKLNLSAGNVDIYKLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLSSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLSSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Qy 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTNDIALMKLOKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTNDIALMKLOKPLTFNDL 360
Qy 361 VKPVCPLNPGMMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVCPLNPGMMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
Qy 421 TPAMICAGFLQGNVDSQCGDGGPLVTSKNINWLLIGDTSWGSCKAKAYRPGVYGNVMVF 480
Db 421 TPAMICAGFLQGNVDSQCGDGGPLVTSKNINWLLIGDTSWGSCKAKAYRPGVYGNVMVF 480
Qy 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

RESULT 4
US-09-780-669-932
Sequence 932, Application US/09780669
Patent No. US20020051977A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqui

APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C24
CURRENT APPLICATION NUMBER: US/09/780,669
CURRENT FILING DATE: 2001-02-09
NUMBER OF SEQ ID NOS: 943
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
TYPE: PRT
LENGTH: 492
ORGANISM: Homo sapiens
US-09-780-669-932

Query Match 99.7%; Score 2709; DB 9; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MALNSGSPALPGYENHGYQENPYPAQPTVTVTVYEVHPAQYYPSPVQYAPRVLTQA 60
Db 1 MALNSGSPALPGYENHGYQENPYPAQPTVTVTVYEVHPAQYYPSPVQYAPRVLTQA 60
Qy 61 SNPVVCTOPKSPSGTCTKTKALCITLTGTLVGAALAAAGLLWKFMSKCSNSGIEC 120
Db 61 SNPVVCTOPKSPSGTCTKTKALCITLTGTLVGAALAAAGLLWKFMSKCSNSGIEC 120
Qy 121 DSSGTCINPSNWCDCVSHCPGGEDENRCVRLYGNFILOVYSSORKSWHPVCQDDWNEY 180
Db 121 DSSGTCINPSNWCDCVSHCPGGEDENRCVRLYGNFILOVYSSORKSWHPVCQDDWNEY 180
Qy 181 GRAACRDMGYKNFYSSQGIIVDDSGTSFMKLNLSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGIIVDDSGTSFMKLNLSAGNVDIYKLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLSSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLSSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Qy 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTNDIALMKLOKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTNDIALMKLOKPLTFNDL 360
Qy 361 VKPVCPLNPGMMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVCPLNPGMMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
Qy 421 TPAMICAGFLQGNVDSQCGDGGPLVTSKNINWLLIGDTSWGSCKAKAYRPGVYGNVMVF 480
Db 421 TPAMICAGFLQGNVDSQCGDGGPLVTSKNINWLLIGDTSWGSCKAKAYRPGVYGNVMVF 480
Qy 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492

RESULT 5
US-09-822-827-932
Sequence 932, Application US/09822827
Patent No. US20020081680A1

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-822-827-932

Query Match 99.7%; Score 2709; DB 9; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHPAQQYPSVPOYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHPAQQYPSVPOYAPRVLTQA 60
QY 61 SNPVCTOPKPSGTVCTSKTKKALCITLTGTFVLGAAALAGLLWKFMSKCSNSGIEC 120
Db 61 SNPVCTOPKPSGTVCTSKTKKALCITLTGTFVLGAAALAGLLWKFMSKCSNSGIEC 120
QY 121 DSSGTCINPNCWCDGVSHCPGEGEDENRCVRLYGNPFIQVYSSQKSHHPVCQDDWNEY 180
Db 121 DSSGTCINPNCWCDGVSHCPGEGEDENRCVRLYGNPFIQVYSSQKSHHPVCQDDWNEY 180
QY 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGSGESALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIVGSGESALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCPLPNPQWMLQPEQLCWISGWGATEKGTSEVLNAAKVLIETQRCNSRYVDNLI 420
Db 361 VKPVCPLPNPQWMLQPEQLCWISGWGATEKGTSEVLNAAKVLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLOQNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGYGVNVWF 480
Db 421 TPAMICAGFLOQNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGYGVNVWF 480
QY 481 TDWYRQMRADG 492
Db 481 TDWYRQMRADG 492

RESULT 6

US-09-895-793-932
; Sequence 932, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.

RESULT 7

US-09-895-814-932
; Sequence 932, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.

; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Basols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-932

Query Match 99.7%; Score 2709; DB 9; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHPAQQYPSVPOYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHPAQQYPSVPOYAPRVLTQA 60
QY 61 SNPVCTOPKPSGTVCTSKTKKALCITLTGTFVLGAAALAGLLWKFMSKCSNSGIEC 120
Db 61 SNPVCTOPKPSGTVCTSKTKKALCITLTGTFVLGAAALAGLLWKFMSKCSNSGIEC 120
QY 121 DSSGTCINPNCWCDGVSHCPGEGEDENRCVRLYGNPFIQVYSSQKSHHPVCQDDWNEY 180
Db 121 DSSGTCINPNCWCDGVSHCPGEGEDENRCVRLYGNPFIQVYSSQKSHHPVCQDDWNEY 180
QY 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGSGESALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIVGSGESALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVIHSPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCPLPNPQWMLQPEQLCWISGWGATEKGTSEVLNAAKVLIETQRCNSRYVDNLI 420
Db 361 VKPVCPLPNPQWMLQPEQLCWISGWGATEKGTSEVLNAAKVLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLOQNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGYGVNVWF 480
Db 421 TPAMICAGFLOQNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGYGVNVWF 480
QY 481 TDWYRQMRADG 492
Db 481 TDWYRQMRADG 492

; APPLICANT: Jiang, Yudi
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedwick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Hural, John
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Vinals de Bassols, Carlota
 ; APPLICANT: Foy, Teresa
 ; APPLICANT: Fanger, Gary R.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 ; FILE REFERENCE: 210121.427C26
 ; CURRENT APPLICATION NUMBER: US/09/895,814
 ; CURRENT FILING DATE: 2001-06-29
 ; NUMBER OF SEQ ID NOS: 990
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 932
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-895-814-932

Query Match 99.7%; Score 2709; DB 9; Length 492;
 Best Local Similarity 99.8%; Pred. No. 2.6e-229;
 Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MALNSGSPPAIGPYENHGYOPENPYPAQPTVPTVVEVHPAQYPSVPQYAPRVLTQA	60
Db	1	MALNSGSPPAIGPYENHGYOPENPYPAQPTVPTVVEVHPAQYPSVPQYAPRVLTQA	60
QY	61	SNPVCTOPKSPSGVCTSKTKKALCITLTGTLVGAALAGLLWKFMSKCSNSGIEC	120
Db	61	SNPVCTOPKSPSGVCTSKTKKALCITLTGTLVGAALAGLLWKFMSKCSNSGIEC	120
QY	121	DSSGTCINPNCWCDGVSHCPGDEENRCVRLYGPNFLOVYSSORKSWHPVCDDMNENY	180
Db	121	DSSGTCINPNCWCDGVSHCPGDEENRCVRLYGPNFLOVYSSORKSWHPVCDDMNENY	180
QY	181	GRAACRDWGYKKNFYSSOGIIVDDSGTSMKLNLSAGNVDIYKLYHSDACSKAVVSLR	240
Db	181	GRAACRDWGYKKNFYSSOGIIVDDSGTSMKLNLSAGNVDIYKLYHSDACSKAVVSLR	240
QY	241	CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK	300
Db	241	CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK	300
QY	301	PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTNNDIAMKLOKPLTFNDL	360
Db	301	PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTNNDIAMKLOKPLTFNDL	360
QY	361	VKPVCLPNPMMLOPEQLCWISGSGATEEKGKTSVLNAKVLIIETQRCNSRYVDNLI	420
Db	361	VKPVCLPNPMMLOPEQLCWISGSGATEEKGKTSVLNAKVLIIETQRCNSRYVDNLI	420
QY	421	TPAMICAGFLOQVDSOCGSGPLVTSKNNIWLIGDTSWGSCKAYRPGYGVNMVF	480
Db	421	TPAMICAGFLOQVDSOCGSGPLVTSKNNIWLIGDTSWGSCKAYRPGYGVNMVF	480
QY	481	TDWIYRQMRADG	492
Db	481	TDWIYRQMRADG	492

RESULT 8

US-10-012-896-932
 ; Sequence 932, Application US/10012896
 ; Publication No. US20020183251A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yudi
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedwick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Hural, John
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Vinals de Bassols, Carlota
 ; APPLICANT: Foy, Teresa
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Mantanabe, Yoshihiro
 ; APPLICANT: Meagher, Madeleine Joy
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 ; FILE REFERENCE: 210121.427C27
 ; CURRENT APPLICATION NUMBER: US/10/012,896
 ; CURRENT FILING DATE: 2001-12-10
 ; NUMBER OF SEQ ID NOS: 1011
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 932
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-012-896-932

Query Match 99.7%; Score 2709; DB 13; Length 492;
 Best Local Similarity 99.8%; Pred. No. 2.6e-229;
 Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MALNSGSPPAIGPYENHGYOPENPYPAQPTVPTVVEVHPAQYPSVPQYAPRVLTQA	60
Db	1	MALNSGSPPAIGPYENHGYOPENPYPAQPTVPTVVEVHPAQYPSVPQYAPRVLTQA	60
QY	61	SNPVCTOPKSPSGVCTSKTKKALCITLTGTLVGAALAGLLWKFMSKCSNSGIEC	120
Db	61	SNPVCTOPKSPSGVCTSKTKKALCITLTGTLVGAALAGLLWKFMSKCSNSGIEC	120
QY	121	DSSGTCINPNCWCDGVSHCPGDEENRCVRLYGPNFLOVYSSORKSWHPVCDDMNENY	180
Db	121	DSSGTCINPNCWCDGVSHCPGDEENRCVRLYGPNFLOVYSSORKSWHPVCDDMNENY	180
QY	181	GRAACRDWGYKKNFYSSOGIIVDDSGTSMKLNLSAGNVDIYKLYHSDACSKAVVSLR	240
Db	181	GRAACRDWGYKKNFYSSOGIIVDDSGTSMKLNLSAGNVDIYKLYHSDACSKAVVSLR	240
QY	241	CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK	300
Db	241	CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK	300
QY	301	PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTNNDIAMKLOKPLTFNDL	360
Db	301	PLNNPWHWTAFAGILRQSFYAGYQVEKVI SHPNYDSKTNNDIAMKLOKPLTFNDL	360
QY	361	VKPVCLPNPMMLOPEQLCWISGSGATEEKGKTSVLNAKVLIIETQRCNSRYVDNLI	420
Db	361	VKPVCLPNPMMLOPEQLCWISGSGATEEKGKTSVLNAKVLIIETQRCNSRYVDNLI	420

QY 421 TPAMICAGFLQGNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVMVF 480
DB 421 TPAMICAGFLQGNVDSQGDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVMVF 480
QY 481 TDWIYRQMRADG 492
DB 481 TDWIYRQMRADG 492

RESULT 9

US-10-144-678A-932
; Sequence 932, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillion, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Steiky, Yasir A. W.
; APPLICANT: Hepler, William T.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals y de Bassols, Carlota
; APPLICANT: Foy, Teresa M.
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Deng, Ta

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.427C28
; CURRENT APPLICATION NUMBER: US/10/144,678A
; CURRENT FILING DATE: 2002-08-12
; NUMBER OF SEQ ID NOS: 1033

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 932
; LENGTH: 492

; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-144-678A-932

Query Match 99.7%; Score 2709; DB 14; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.6e-229;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTOA 60
DB 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTOA 60
QY 61 SNPVVCTOPKSPGTVCTSKTKALCITLTGTLVGAALAGLLWKPMGSKCSNSGIEC 120
DB 61 SNPVVCTOPKSPGTVCTSKTKALCITLTGTLVGAALAGLLWKPMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDCGVSHCPGDEENRCVRLYGPNFLLQVYSSQKSWHPVCQDDWNNY 180
DB 121 DSSGTCINPSNWCDCGVSHCPGDEENRCVRLYGPNFLLQVYSSQKSWHPVCQDDWNNY 180
QY 181 GRAACRDGMGYKNFYSSQGIYDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDGMGYKNFYSSQGIYDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLSSRSQSRIVGGSALPGAWPQVSLHVQNHVHVCVCGSIITPEWIVTAAHCVEK 300
DB 241 CIACGVNLSSRSQSRIVGGSALPGAWPQVSLHVQNHVHVCVCGSIITPEWIVTAAHCVEK 300

Db 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
QY 481 TDWIYRQWRADG 492
Db 481 TDWIYRQWRADG 492

RESULT 11

US-10-334-038-6
; Sequence 6, Application US/10334038
; Publication No. US2003013886A1
; GENERAL INFORMATION:
; APPLICANT: Macbeth, Kyle J.
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NUCLEIC ACIDS AND POLYPEPTIDES USEFUL IN
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF PROSTATE CANCER
; FILE REFERENCE: 07334-135001
; CURRENT APPLICATION NUMBER: US/10/334,038
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US/09/408,628A
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: US 09/164,159
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: US 09/163,759
; PRIOR FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-334-038-6

Query Match 99.5%; Score 2704; DB 14; Length 492;
Best Local Similarity 99.4%; Pred. No. 7.1e-229;
Matches 489; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGSPALGPYYENHGYQENPYPAQPTVPTVVEVHPAQYPSVPQYAPRVLTOA 60
Db 1 MALNSGSPALGPYYENHGYQENPYPAQPTVPTVVEVHPAQYPSVPQYAPRVLTOA 60
QY 61 SNPVVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDCVSHCPGGEDENCVRLYGNFVLQVYSSORKSWHPVCCDDWNNY 180
Db 121 DSSGTCINPSNWCDCVSHCPGGEDENCVRLYGNFVLQVYSSORKSWHPVCCDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGI VDDSGSTSPFKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGI VDDSGSTSPFKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNRRQSRIRVGGESALPGAWPMQVSLHVQNVHVCVCGSIITPEWIVTAACHVEK 300
Db 241 CIACGVNLNRRQSRIRVGGESALPGAWPMQVSLHVQNVHVCVCGSIITPEWIVTAACHVEK 300
QY 301 PLNNPWHWTAFAGILRQSFMEYAGYQVQKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMEYAGYQVQKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCPLNPNQMWLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVCPLNPNQMWLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
Db 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
QY 481 TDWIYRQWRADG 492
Db 481 TDWIYRQWRADG 492

RESULT 12

US-09-825-751A-63
; Sequence 63, Application US/09825751A
; Publication No. US20030065140A1
; GENERAL INFORMATION:
; APPLICANT: Curagen Corporation
; APPLICANT: Vernhet, Corine A.M.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spyttek, Kimberly A.
; APPLICANT: Rastelli, Luca
; APPLICANT: Herltan, John L.
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-750
; CURRENT APPLICATION NUMBER: US/09/825,751A
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/194,314
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/225,693
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-825-751A-63

Query Match 99.4%; Score 2701; DB 12; Length 492;
Best Local Similarity 99.0%; Pred. No. 1.3e-228;
Matches 487; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALNSGSPALGPYYENHGYQENPYPAQPTVPTVVEVHPAQYPSVPQYAPRVLTOA 60
Db 1 MALNSGSPALGPYYENHGYQENPYPAQPTVPTVVEVHPAQYPSVPQYAPRVLTOA 60
QY 61 SNPVVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLWFKMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDCVSHCPGGEDENCVRLYGNFVLQVYSSORKSWHPVCCDDWNNY 180
Db 121 DSSGTCINPSNWCDCVSHCPGGEDENCVRLYGNFVLQVYSSORKSWHPVCCDDWNNY 180
QY 181 GRAACRDMGYKNFYSSQGI VDDSGSTSPFKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGI VDDSGSTSPFKLNTSAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNRRQSRIRVGGESALPGAWPMQVSLHVQNVHVCVCGSIITPEWIVTAACHVEK 300
Db 241 CIACGVNLNRRQSRIRVGGESALPGAWPMQVSLHVQNVHVCVCGSIITPEWIVTAACHVEK 300
QY 301 PLNNPWHWTAFAGILRQSFMEYAGYQVQKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMEYAGYQVQKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCPLNPNQMWLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVCPLNPNQMWLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
Db 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVYGNVWF 480
QY 481 TDWIYRQWRADG 492
Db 481 TDWIYRQWRADG 492

RESULT 13

US-09-759-143-895
; Sequence 895, Application US/09759143
; Patent No. US200202248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759.143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-895

Query Match 99.2%; Score 2696; DB 9; Length 492;
Best Local Similarity 98.8%; Pred. No. 3.6e-228;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHVAQYPSVPVQAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHVAQYPSVPVQAPRVLTQA 60
QY 61 SNPVCTQPKSPSTGCTCTKALCITITLGTFLVGAALAGLLKFKGSKCSNGI 120
Db 61 SNPVCTQPKSPSTGCTCTKALCITITLGTFLVGAALAGLLKFKGSKCSNGI 120
QY 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGPNFLOVYSSQKSWHPVCODDNN 180
Db 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGPNFLOVYSSQKSWHPVCODDNN 180
QY 181 GRAACRDMGYKNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAV 240
Db 181 GRAACRDMGYKNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAV 240
QY 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAH 300
Db 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAH 300
QY 301 PLNNPWHWTAFAGILROSFMFYGAGYQVESHENYDSKTKNDIALMKLOKPLTF 360
Db 301 PLNNPWHWTAFAGILROSFMFYGAGYQVESHENYDSKTKNDIALMKLOKPLTF 360
QY 361 VKPVCLENPQWMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQCN 420
Db 361 VKPVCLENPQWMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQCN 420
QY 421 TPAMICAGFLQGNVDSQCGSGPLVTSNNIWLIGDTSWGGCAKAYRPGYGNV 480
Db 421 TPAMICAGFLQGNVDSQCGSGPLVTSNNIWLIGDTSWGGCAKAYRPGYGNV 480
QY 481 TDWIYROWRADG 492
Db 481 TDWIYROWKANG 492

RESULT 14
US-09-780-669-895
; Sequence 895, Application US/09780669
; Patent No. US20020851977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: McNeill, John
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780.669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-780-669-895

Query Match 99.2%; Score 2696; DB 9; Length 492;
Best Local Similarity 98.8%; Pred. No. 3.6e-228;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHVAQYPSVPVQAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVHVAQYPSVPVQAPRVLTQA 60
QY 61 SNPVCTQPKSPSTGCTCTKALCITITLGTFLVGAALAGLLKFKGSKCSNGI 120
Db 61 SNPVCTQPKSPSTGCTCTKALCITITLGTFLVGAALAGLLKFKGSKCSNGI 120
QY 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGPNFLOVYSSQKSWHPVCODDNN 180
Db 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGPNFLOVYSSQKSWHPVCODDNN 180
QY 181 GRAACRDMGYKNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAV 240
Db 181 GRAACRDMGYKNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAV 240
QY 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAH 300
Db 241 CIACGVNLNSRQSRIVGGSALPGAWPQVSLHVQNVHVCVCGSIITPEWIVTAAH 300
QY 301 PLNNPWHWTAFAGILROSFMFYGAGYQVESHENYDSKTKNDIALMKLOKPLTF 360
Db 301 PLNNPWHWTAFAGILROSFMFYGAGYQVESHENYDSKTKNDIALMKLOKPLTF 360
QY 361 VKPVCLENPQWMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQCN 420
Db 361 VKPVCLENPQWMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQCN 420
QY 421 TPAMICAGFLQGNVDSQCGSGPLVTSNNIWLIGDTSWGGCAKAYRPGYGNV 480

Db 481 TDWIYROMKANG 492
Search completed: June 1, 2004, 14:48:07
Job time : 125 secs

Db 421 TPAMICAGFLQGNVDSQDGGSGPLVTSNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480
Qy 481 TDWIYROMKANG 492
Db 481 TDWIYROMKANG 492

RESULT 15
US-09-879-792-14
; Sequence 14, Application US/09879792
; Patent No. US20020061850A1
; GENERAL INFORMATION:
; APPLICANT: Xiao, Yonghong
; APPLICANT: Gedrich, Richard
; TITLE OF INVENTION: Regulation of Human Transmembrane Serine
; TITLE OF INVENTION: Protease
; FILE REFERENCE: 02973.00035
; CURRENT APPLICATION NUMBER: US/09/879,792
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/211,224
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: US 60/283,353
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/283,648
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: PCT
; PRIOR FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-879-792-14

Query Match 99.2%; Score 2696; DB 9; Length 492;
Best Local Similarity 98.8%; Pred. No. 3.6e-228;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVYVHPAQYYPSPVQYAPRVLTOA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVYVHPAQYYPSPVQYAPRVLTOA 60
Qy 61 SNPVVCTOPKPSGTVCTSKTKALCITLTGTLVGAALAAAGLLKFMGSKCSNSGIEC 120
Db 61 SNPVVCTOPKPSGTVCTSKTKALCITLTGTLVGAALAAAGLLKFMGSKCSNSGIEC 120
Qy 121 DSSGTCINPSNWCDCVSHCPGSEDENRCVLYGNPFIQVYSSORKSWHPVCQDDWNEY 180
Db 121 DSSGTCINPSNWCDCVSHCPGSEDENRCVLYGNPFIQVYSSORKSWHPVCQDDWNEY 180
Qy 181 GRAACRDMGYKKNFYSSQGIYVDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
Db 181 GRAACRDMGYKKNFYSSQGIYVDSGSTSFMKLNTSAGNVDIYKLYHSDACSKAVVSLR 240
Qy 241 CIACGVNLNSRQSRIVGESALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIVGESALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
Qy 301 PLNNPWHMTAFAGILRQSPMFYAGYQVEKVIHSHENYDSKYNNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHMTAFAGILRQSPMFYAGYQVEKVIHSHENYDSKYNNDIALMKLQKPLTFNDL 360
Qy 361 VKPVCLPNPQMLQBPQLCWIISGWGATEBKGTSEVLNAKVLIIETQRCNSRYVDNLI 420
Db 361 VKPVCLPNPQMLQBPQLCWIISGWGATEBKGTSEVLNAKVLIIETQRCNSRYVDNLI 420
Qy 421 TPAMICAGFLQGNVDSQDGGSGPLVTSNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480
Db 421 TPAMICAGFLQGNVDSQDGGSGPLVTSNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480
Qy 481 TDWIYROMKANG 492

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 14:04:23 ; Search time 136 Seconds
(without alignments)
1022.157 Million cell updates/sec

Title: US-09-615-285B-2
Perfect score: 217
Sequence: 1 MALNSGSPPAIGPYVENHGY.....VIGNVMVFTDIYQMRADG 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A Geneseq 29Jan04:*
- 1: geneseqp1980s:*
 - 2: geneseqp1990s:*
 - 3: geneseqp2000s:*
 - 4: geneseqp2001s:*
 - 5: geneseqp2002s:*
 - 6: geneseqp2003as:*
 - 7: geneseqp2003bs:*
 - 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	492	AA44406	Aay44406 Human 20P
2	2717	100.0	492	AA77726	Aay77726 Human tum
3	2717	100.0	492	AAE18096	Aae18096 Human 20P
4	2714	99.9	492	AAE18098	Aae18098 Human 20P
5	2713	99.9	492	AAE18100	Aae18100 Human 20P
6	2713	99.9	492	AAE18099	Aae18099 Human 20P
7	2713	99.9	492	AAE18101	Aae18101 Human 20P
8	2711	99.8	492	ADB75591	Adb75591 Prostate
9	2709	99.7	492	AA92050	AA92050 HrpCa6/7
10	2709	99.7	492	AAAM01315	AAm01315 P1000C am
11	2709	99.7	492	AAU69960	Aau69960 Human pro
12	2708	99.7	492	ABU71851	Abu71851 Prostate
13	2709	99.7	492	ABBS5420	Abbs5420 Human p10
14	2709	99.7	492	ABR54532	ABr54532 Prostate
15	2709	99.7	492	ADBI4382	Adbi4382 Human pro
16	2696	99.2	492	AA57280	AA57280 Ovr115 ho
17	2696	99.2	492	AA36901	AA36901 Human TMP
18	2696	99.2	492	AAAM01294	AAm01294 Human tra
19	2696	99.2	492	AAU69939	Aau69939 Human tra
20	2696	99.2	492	AAE06943	Aae06943 Human ser
21	2696	99.2	492	ABU71830	Abu71830 Prostate
22	2696	99.2	492	ABBS95399	Abbs95399 Human tra
23	2696	99.2	492	ABGG61885	ABg61885 Prostate
24	2696	99.2	492	AAE18097	Aae18097 Human TMP
25	2696	99.2	492	ABR48241	ABr48241 Human bla

26	2696	99.2	492	6	ABR54511	ABr54511 Human tra
27	2696	99.2	492	7	ADBI4345	Adbi4345 Human tra
28	2171	79.9	393	4	AAAM01317	AAm01317 P1000C pa
29	2171	79.9	393	4	AAU69962	Aau69962 Human pro
30	2171	79.9	393	4	ABU71853	Abu71853 Prostate
31	2171	79.9	393	5	ABBS5422	Abbs5422 Human p10
32	2171	79.9	393	5	ABR54534	ABr54534 Prostate
33	2171	79.9	393	7	ADBI4384	Adbi4384 Human pro
34	2126	78.2	384	6	ADAI0905	Ada10905 Human CDN
35	1540	56.7	283	3	AA578328	AA578328 Amino aci
36	1540	56.7	283	5	AAE14615	Aae14615 Human tra
37	1165	42.9	209	4	AAAM01295	AAm01295 Human tra
38	1165	42.9	209	4	AAU69940	Aau69940 Human tra
39	1165	42.9	209	4	ABU71831	Abu71831 Prostate
40	1165	42.9	209	5	ABBS5400	Abbs5400 Human tra
41	1165	42.9	209	6	ABR54512	ABr54512 Partial H
42	1165	42.9	209	7	ADBI4347	Adbi4347 Human tra
43	1019.5	37.5	418	6	ABP56845	ABp56845 Consensus
44	888	32.7	454	3	AB332246	AB332246 Tumour as
45						

ALIGNMENTS

RESULT 1
AA44406
ID AAY44406 standard; protein; 492 AA.
XX AC AAY44406;
XX XX
DT 22-MAR-2000 (first entry)
XX XX
DE Human 20P1F12-GTC2 protein.
XX XX
KW 20P1F12; TMPRSS2; androgen; serine protease; 20P1F12-GTC1; cancer;
transmembrane protein; colon; prostate; prostate tumour.
XX OS Homo sapiens.
XX XX
PN WO9962942-A2.
XX XX
PD 09-DEC-1999.
XX XX
PF 01-JUN-1999; 99WO-US012253.
XX XX
PR 01-JUN-1998; 98US-0087598P.
PR 29-JUN-1998; 98US-0091474P.
PR 14-APR-1999; 99US-0129521P.
XX XX
(UROC-) UROGENESYS INC.
PA (AFAR/) AFAR D E.
PA (HUBE/) HUBERT R S.
PA (LEON/) LEONG K.
PA (RAIT/) RAITANO A B.
PA (SAFF/) SAFFRAN D C.
XX XX
PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;
XX XX
DR WPI; 2000-116363/10.
XX XX
N-PSDB; AAZ29636.
PT Novel cell surface antigen useful to treat colon and prostate cancer.
XX XX
PS Claim 1; Fig 1; 58pp; English.
XX XX
CC The present sequence is the 20P1F12 protein (also known as the TMPRSS2
protein) which is a prostate-specific, androgen-regulated, cell surface
serine protease. It is a glycosylated type II transmembrane protein with
an extracellular C-terminal serine protease domain, a scavenger receptor
cysteine-rich domain, an LDL receptor class A domain and a predicted
transmembrane domain. Host cells can be transformed to produce this
protein, using vector containing 20P1F12/TMPRSS2 gene (also designated

CC 20P1F12-GTC1, as deposited with ATCC accession number 207097). Anti-
CC 20P1F12/TMPRSS2 antibodies may be used as therapeutic agent for prostate
CC and colon cancers, to image prostate cancer cells and prostate tumours,
CC to identify ligands and cellular constituents that bind to a
CC 20P1F12/TMPRSS2 gene product and for use as cancer vaccines
XX
XX
SQ Sequence 492 AA;

Query Match 100.0%; Score 2717; DB 3; Length 492;
Best Local Similarity 100.0%; Pred. No. 6.7e-203;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTQA 60
DB 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTQA 60
QY 61 SNPVCTOPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLKFWGSKCSNGIEC 120
DB 61 SNPVCTOPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLKFWGSKCSNGIEC 120
QY 121 DSSGTCINPSNWCDSVGHCPGDEENRCVLYGNFVLQVYSSQKSWHPVCQDDWNNY 180
DB 121 DSSGTCINPSNWCDSVGHCPGDEENRCVLYGNFVLQVYSSQKSWHPVCQDDWNNY 180
QY 181 GRAACRDWGYKNFYSSQGIYVDSGTSFMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDWGYKNFYSSQGIYVDSGTSFMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGGSALPGAMPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
DB 241 CIACGVNLNSRQSRIVGGSALPGAMPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
DB 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCLEPNPGMWLOPEQLCWISGWGATEEGKTSVLNAAKVLIIETQRCNSRYVDNLI 420
DB 361 VKPVCLEPNPGMWLOPEQLCWISGWGATEEGKTSVLNAAKVLIIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLOGNVDSQSGSGPLVTSKNINWLLIGDTSWGSCKAKAYRPGVGNVMVF 480
DB 421 TPAMICAGFLOGNVDSQSGSGPLVTSKNINWLLIGDTSWGSCKAKAYRPGVGNVMVF 480
QY 481 TDWYIYRQMRADG 492
DB 481 TDWYIYRQMRADG 492

RESULT 2

ID AAY77726 standard; protein; 492 AA.

AC AAY77726;

DT 12-MAY-2000 (first entry)

DE Human tumour suppressor TMPRSS2 polypeptide.

XX Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
KW gene therapy; protein therapy.

OS Homo sapiens.

FN WO200000605-A1.

PD 06-JAN-2000,

PF 29-JUN-1999; 99WO-US014622.

PR 29-JUN-1998; 98US-0091044P.

XX (MYRI-) MYRIAD GENETICS INC.

XX Wong AKC, Tavtigian SV, Teng DHP;
XX WPI; 2000-170914/15.
DR N-PSDB; AA287786.
XX Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of
XX human cancer.
XX Claim 55; Page 77-79; 89pp; English.
XX The invention provides a new tumour suppressor gene, designated TMPRSS2.
XX The TMPRSS2 polynucleotides and polypeptides can be used in methods for
XX diagnosing and prognosing predisposition to cancer in humans. The
XX polypeptides may also be used in assays to screen for compounds with anti-
XX cancer or therapeutic properties. The polypeptides are also useful for
XX rational drug design. The TMPRSS2 polynucleotides and polypeptides may be
XX used for gene therapy and protein therapy. The present sequence
XX represents the TMPRSS2 polypeptide
XX
SQ Sequence 492 AA;

Query Match 100.0%; Score 2717; DB 3; Length 492;

Best Local Similarity 100.0%; Pred. No. 6.7e-203;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTQA 60
DB 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVPTVVEVHPAQYYPSPVQYAPRVLTQA 60
QY 61 SNPVCTOPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLKFWGSKCSNGIEC 120
DB 61 SNPVCTOPKSPSGTCTSKTKKALCITLTGLTFLVGAALAAAGLLKFWGSKCSNGIEC 120
QY 121 DSSGTCINPSNWCDSVGHCPGDEENRCVLYGNFVLQVYSSQKSWHPVCQDDWNNY 180
DB 121 DSSGTCINPSNWCDSVGHCPGDEENRCVLYGNFVLQVYSSQKSWHPVCQDDWNNY 180
QY 181 GRAACRDWGYKNFYSSQGIYVDSGTSFMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDWGYKNFYSSQGIYVDSGTSFMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGGSALPGAMPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
DB 241 CIACGVNLNSRQSRIVGGSALPGAMPQVSLHVQNVHVCVCGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
DB 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCLEPNPGMWLOPEQLCWISGWGATEEGKTSVLNAAKVLIIETQRCNSRYVDNLI 420
DB 361 VKPVCLEPNPGMWLOPEQLCWISGWGATEEGKTSVLNAAKVLIIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLOGNVDSQSGSGPLVTSKNINWLLIGDTSWGSCKAKAYRPGVGNVMVF 480
DB 421 TPAMICAGFLOGNVDSQSGSGPLVTSKNINWLLIGDTSWGSCKAKAYRPGVGNVMVF 480
QY 481 TDWYIYRQMRADG 492
DB 481 TDWYIYRQMRADG 492

RESULT 3

AAE18096

ID AAE18096 standard; protein; 492 AA.

XX AAE18096;

XX 07-MAY-2002 (first entry)

XX Human 20P1F12-GTC1 protein.

KW	Serine protease; 20PiF12/TMPRSS2; 20PiF12-GTC1; cell growth; neoplasm;	
KW	cancer; vaccine; human.	
XX		
OS	Homo sapiens.	
XX		
FN	WO200204953-A2.	
XX		
PD	17-JAN-2002.	
XX		
PF	12-JUL-2001; 2001WO-US022168.	
XX		
PR	12-JUL-2000; 2000US-00615285.	
XX		
PA	(AGEN-) AGENSYS INC.	
XX		
PI	Saferan D, Raitano AB, Hubert RS, Jakobovits A, Faris M;	
PI	Challita-Eid PM;	
XX		
DR	WPI; 2002-154967/20.	
XX		
DR	N-PSDB; AAD28778.	
XX		
PT	Examining a biological sample for evidence of dysregulated cellular	
PT	growth, comprises comparing the status of prostate-specific, androgen-	
PT	regulated, secreted serine protease, 20PiF12/TMPRSS2, in a corresponding	
PT	normal sample.	
XX		
PS	Claim 40; Fig 1; 161pp; English.	
XX		
CC	The present invention relates to methods and compositions for the	
CC	diagnosis and therapy of prostate, colon, bladder, lung, ovarian and	
CC	kidney cancer derived from or based on a normally prostate-specific,	
CC	androgen regulated, cell membrane associated secreted serine protease	
CC	termed 20PiF12/TMPRSS2. The invention further relates to a method of	
CC	examining a biological sample for evidence of dysregulated cellular	
CC	growth comprises comparing the status of 20PiF12/TMPRSS2 gene (also	
CC	designated 20PiF12-GTC1) in the sample to the status of 20PiF12/TMPRSS2	
CC	in a corresponding normal sample. The invention also relates to 20PiF12/	
CC	TMPRSS2 polynucleotides and their corresponding proteins. Methods of the	
CC	invention are used for examining a sample such as blood, serum, stool,	
CC	urine, semen, or biopsy tissue for evidence of dysregulated cell growth.	
CC	The dysregulated cell growth is indicative of bladder cancer, lung	
CC	cancer, kidney cancer or ovarian cancer. It is useful for identifying	
CC	evidence of a neoplasm in a sample. Vaccines comprising an immunogenic	
CC	portion of 20PiF12/TMPRSS2 are useful for inhibiting growth of a cell	
CC	expressing 20PiF12/TMPRSS2 in a patient suffering from bladder cancer,	
CC	lung cancer, ovarian cancer or metastatic cancer. The present sequence	
CC	human 20PiF12-GTC1 protein	
XX		
SQ	Sequence 492 AA;	
	Query Match 100.0%; Score 2717; DB 5; Length 492;	
	Best Local Similarity 100.0%; Pred. No. 6.7e-203; Mismatches 0; Gaps 0;	
	Matches 492; Conservative 0; Indels 0; Gaps 0;	
QY	1 MALNSGSPPAIGPYENHGYENHPYPAQPTVPTVYEVHPAQYVPSVPQYAPVLTQA 60	
DB	1 MALNSGSPPAIGPYENHGYENHPYPAQPTVPTVYEVHPAQYVPSVPQYAPVLTQA 60	
QY	61 SNPVVCTQKSPSGVCTSKTKKALCITLTGTLVGAALAGLLKFMGSKNSGIEC 120	
DB	61 SNPVVCTQKSPSGVCTSKTKKALCITLTGTLVGAALAGLLKFMGSKNSGIEC 120	
QY	121 DSSGTCINPSNWDGSHCPGGEENRCVRLYGNPFIQVYSSQKSHHPVQDDWNEY 180	
DB	121 DSSGTCINPSNWDGSHCPGGEENRCVRLYGNPFIQVYSSQKSHHPVQDDWNEY 180	
QY	181 GRAACRDGMGYKXNFYSSQGVVDDSGSTFPMKLNISAGNVDIYKLYHSDACSSKAVVSLR 240	
DB	181 GRAACRDGMGYKXNFYSSQGVVDDSGSTFPMKLNISAGNVDIYKLYHSDACSSKAVVSLR 240	
QY	241 CIACGVNLNSRQSRIVGESALPGAWPQVSLHVONVHVCVGSIIIPETWITVTAHCVCK 300	
DB	241 CIACGVNLNSRQSRIVGESALPGAWPQVSLHVONVHVCVGSIIIPETWITVTAHCVCK 300	

QY	301 PLNNPHWHTAFAGILRQSPFMFYGAGYQVEKVIKSHPNYDSKTKNNDIALMKLOKPLTFNDL 360	
DB	301 PLNNPHWHTAFAGILRQSPFMFYGAGYQVEKVIKSHPNYDSKTKNNDIALMKLOKPLTFNDL 360	
QY	361 VKPVCILPNPGMMLQPEQLCWISGWGATEBKGTSEVLNAKVLIIETQRCNSRYVDNLI 420	
DB	361 VKPVCILPNPGMMLQPEQLCWISGWGATEBKGTSEVLNAKVLIIETQRCNSRYVDNLI 420	
QY	421 TPAMICAGFLOQNVDSQCGDGGPLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480	
DB	421 TPAMICAGFLOQNVDSQCGDGGPLVTSKNNIWWLIGDTSWGGCAKAYRPGVYGNVMVF 480	
QY	481 TDWIYRQMRADG 492	
DB	481 TDWIYRQMRADG 492	
RESULT 4		
AAE18098	ID	AAE18098 standard; protein; 492 AA.
XX	AC	AAE18098;
XX	DT	07-MAY-2002 (first entry)
XX	XX	Human 20PiF12/TMPRSS2 mutant protein (S441A).
XX	DE	Serine protease; 20PiF12/TMPRSS2; 20PiF12-GTC1; cell growth; neoplasm;
KW	KW	cancer; vaccine; human.
XX	XX	Homo sapiens.
OS	OS	Synthetic.
XX	XX	Key
XX	XX	Location/Qualifiers
FT	FT	Misc-difference 441
FT	FT	/note= "Wild type Ser substituted with Ala"
XX	XX	WO200204953-A2.
XX	XX	17-JAN-2002.
XX	XX	12-JUL-2001; 2001WO-US022168.
XX	XX	12-JUL-2000; 2000US-00615285.
XX	XX	(AGEN-) AGENSYS INC.
XX	XX	Saferan D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
XX	XX	Challita-Eid PM;
XX	XX	WPI; 2002-154967/20.
XX	XX	Examining a biological sample for evidence of dysregulated cellular
XX	XX	growth, comprises comparing the status of prostate-specific, androgen-
XX	XX	regulated, secreted serine protease, 20PiF12/TMPRSS2, in a corresponding
XX	XX	normal sample.
XX	XX	Example 10; Page; 161pp; English.
XX	XX	The present invention relates to methods and compositions for the
XX	XX	diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX	XX	kidney cancer derived from or based on a normally prostate-specific,
XX	XX	androgen regulated, cell membrane associated secreted serine protease
XX	XX	termed 20PiF12/TMPRSS2. The invention further relates to a method of
XX	XX	examining a biological sample for evidence of dysregulated cellular
XX	XX	growth comprises comparing the status of 20PiF12/TMPRSS2 gene (also
XX	XX	designated 20PiF12-GTC1) in the sample to the status of 20PiF12/TMPRSS2
XX	XX	in a corresponding normal sample. The invention also relates to 20PiF12/
XX	XX	TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
XX	XX	invention are used for examining a sample such as blood, serum, stool,
XX	XX	urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX	XX	The dysregulated cell growth is indicative of bladder cancer, lung
XX	XX	cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX	XX	evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX	XX	portion of 20PiF12/TMPRSS2 are useful for inhibiting growth of a cell
XX	XX	expressing 20PiF12/TMPRSS2 in a patient suffering from bladder cancer,
XX	XX	lung cancer, ovarian cancer or metastatic cancer. The present sequence
XX	XX	human 20PiF12-GTC1 protein

CC cancer, kidney cancer or ovarian cancer. It is useful for identifying
CC evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
CC portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
CC expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
CC lung cancer, ovarian cancer or metastatic cancer. The present sequence is
CC human 20P1F12/TMPRSS2 mutant protein (S441A). Note: This sequence is not
CC shown in the specification but is derived from wild type 20P1F12/TMPRSS2
CC protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX
SQ Sequence 492 AA;

Query Match 99.9%; Score 2714; DB 5; Length 492;
Best Local Similarity 99.8%; Pred. No. 1.4e-202;
Matches 491; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVYEVHPAQYYPSPVQYAPRVLTOA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVYEVHPAQYYPSPVQYAPRVLTOA 60

Qy 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNGIEC 120
Db 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNGIEC 120

Qy 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFLLQVYSSQKSWHPVCQDDWNNY 180
Db 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFLLQVYSSQKSWHPVCQDDWNNY 180

Qy 181 GRAACRDMGYKNNFYSSQGIIVDDSGTSPFMKLTNSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNNFYSSQGIIVDDSGTSPFMKLTNSAGNVDIYKLYHSDACSSKAVVSLR 240

Qy 241 CIACGVNLNSRQSRIRVGGESALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIRVGGESALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300

Qy 301 PLNNPWHWTAPAGILRQSFYAGYQVEKVI SHPNYDSTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAPAGILRQSFYAGYQVEKVI SHPNYDSTKNDIALMKLQKPLTFNDL 360

Qy 361 VKPVCLENPGMWLOPEQLCISGWGATEEKGKTSVLNAAKVLIIETQRCNSRYVDNLI 420
Db 361 VKPVCLENPGMWLOPEQLCISGWGATEEKGKTSVLNAAKVLIIETQRCNSRYVDNLI 420

Qy 421 TPAMICAGFLOGNVDSCQSGSGLPLVTSKNNIWLIGDTSWGGCAKAYRPGYGNVMVF 480
Db 421 TPAMICAGFLOGNVDSCQSGSGLPLVTSKNNIWLIGDTSWGGCAKAYRPGYGNVMVF 480

Qy 481 TDWIYRQNRADG 492
Db 481 TDWIYRQNRADG 492

RESULT 5
AAE18100
ID AAE18100 standard; protein; 492 AA.
XX
AC AAE18100;
XX
XX
DT 07-MAY-2002 (first entry)
XX
DE Human 20P1F12/TMPRSS2 mutant protein (R252Q).
XX
XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTC1; cell growth; neoplasm;
KW cancer; vaccine; human.
XX
XX Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 252
XX /note= "wild type Arg substituted with Gln"
XX
PN WO200204953-A2.

XX 17-JAN-2002.
XX 12-JUL-2001; 2001WO-US022168.
XX 12-JUL-2000; 2000US-00615285.
XX (AGEN-) AGENSYS INC.
XX
XX Saferan D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
XX Challita-Eid PM;
XX WPI; 2002-154967/20.
XX
XX Examining a biological sample for evidence of dysregulated cellular
XX growth, comprises comparing the status of prostate-specific, androgen-
XX regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
XX normal sample.
XX
XX Example 10; Page; 161pp; English.
XX
XX The present invention relates to methods and compositions for the
XX diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX kidney cancer derived from or based on a normally prostate-specific,
XX androgen regulated, cell membrane associated secreted serine protease
XX termed 20P1F12/TMPRSS2. The invention further relates to a method of
XX examining a biological sample for evidence of dysregulated cellular
XX growth comprises comparing the status of 20P1F12/TMPRSS2 gene (also
XX designated 20P1F12-GTC1) in the sample to the status of 20P1F12/TMPRSS2
XX in a corresponding normal sample. The invention also relates to 20P1F12/
XX TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
XX invention are used for examining a sample such as blood, serum, stool,
XX urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX The dysregulated cell growth is indicative of bladder cancer, lung
XX cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
XX expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
XX lung cancer, ovarian cancer or metastatic cancer. The present sequence is
XX human 20P1F12/TMPRSS2 mutant protein (R252Q). Note: This sequence is not
XX shown in the specification but is derived from wild type 20P1F12/TMPRSS2
XX protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX
SQ Sequence 492 AA;

Query Match 99.9%; Score 2713; DB 5; Length 492;
Best Local Similarity 99.8%; Pred. No. 1.4e-202;
Matches 491; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVYEVHPAQYYPSPVQYAPRVLTOA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVVYEVHPAQYYPSPVQYAPRVLTOA 60

Qy 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNGIEC 120
Db 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLKFMGSKCSNGIEC 120

Qy 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFLLQVYSSQKSWHPVCQDDWNNY 180
Db 121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFLLQVYSSQKSWHPVCQDDWNNY 180

Qy 181 GRAACRDMGYKNNFYSSQGIIVDDSGTSPFMKLTNSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNNFYSSQGIIVDDSGTSPFMKLTNSAGNVDIYKLYHSDACSSKAVVSLR 240

Qy 241 CIACGVNLNSRQSRIRVGGESALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIRVGGESALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300

Qy 301 PLNNPWHWTAPAGILRQSFYAGYQVEKVI SHPNYDSTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAPAGILRQSFYAGYQVEKVI SHPNYDSTKNDIALMKLQKPLTFNDL 360

QY 361 VKFVCLPDPGMMLOPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKFVCLPDPGMMLOPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSGQSGPLVTSKNNIWMILIGDTSWGSCKAKAYRPGVGNMVF 480
Db 421 TPAMICAGFLQGNVDSGQSGPLVTSKNNIWMILIGDTSWGSCKAKAYRPGVGNMVF 480
QY 481 TDWIYQMRADG 492
Db 481 TDWIYQMRADG 492

RESULT 6
AAE18099
ID AAE18099 standard; protein; 492 AA.
XX AAE18099;
XX
XX 07-MAY-2002 (first entry)
XX Human 20P1F12/TMPRSS2 mutant protein (R240Q).
XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCl; cell growth; neoplasm;
KW cancer; vaccine; human.
XX Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 240 /note= "Wild type Arg substituted with Gln"
FT
XX WO200204953-A2.
XX 17-JAN-2002.
XX 12-JUL-2001; 2001WO-US022168.
XX 12-JUL-2000; 2000US-00615285.
XX (AGEN-) AGENSYS INC.
XX Saferran D, Raitano AB, Hubert RS, Jakobovits A, Paris M;
PI Challita-Eid PM;
XX WPI; 2002-154967/20.
XX
XX Examining a biological sample for evidence of dysregulated cellular
PT growth, comprises comparing the status of prostate-specific, androgen-
PT regulated, secreted serine protease, 20P1F12/TMPRSS2, in a corresponding
PT normal sample.
XX
XX Example 10; Page; 161pp; English.
XX
XX The present invention relates to methods and compositions for the
CC diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
CC kidney cancer derived from or based on a normally prostate-specific,
CC androgen regulated, cell membrane associated secreted serine protease
CC termed 20P1F12/TMPRSS2. The invention further relates to a method of
CC examining a biological sample for evidence of dysregulated cellular
CC growth comprising comparing the status of 20P1F12/TMPRSS2 gene (also
CC designated 20P1F12-GTCl) in the sample to the status of 20P1F12/
CC in a corresponding normal sample. The invention also relates to 20P1F12/
CC TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
CC invention are used for examining a sample such as blood, serum, stool,
CC urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
CC The dysregulated cell growth is indicative of evidence of bladder cancer, lung
CC cancer, kidney cancer or ovarian cancer. It is useful for identifying
CC evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
CC portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
CC expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
CC lung cancer, ovarian cancer or metastatic cancer. The present sequence is

CC human 20P1F12/TMPRSS2 mutant protein (R240Q). Note: This sequence is not
CC shown in the specification but is derived from wild type 20P1F12/TMPRSS2
CC protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX
SQ Sequence 492 AA;
Query Match 99.9%; Score 2713; DB 5; Length 492;
Best Local Similarity 99.8%; Pred. No. 1.4e-202; Indels 0; Gaps 0;
Matches 491; Conservative 1; Mismatches 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVVPTVYEVHFAQYFSPVPQYAPRVLTA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVVPTVYEVHFAQYFSPVPQYAPRVLTA 60
QY 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLNKFWSKCSNSGIEC 120
Db 61 SNPVCTQPKSPSGTCTSKTKKALCITLTGLTFLVGAALAGLLNKFWSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDSVSHCPGGEDENKCVRLYGFNFILQVYSSQKSWHPVCCDDWNEY 180
Db 121 DSSGTCINPSNWCDSVSHCPGGEDENKCVRLYGFNFILQVYSSQKSWHPVCCDDWNEY 180
QY 181 GRAACRDMGYKNNFYSSQGIIVDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLQ 240
Db 181 GRAACRDMGYKNNFYSSQGIIVDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVVSLQ 240
QY 241 CIACGVNLSSRSQSRIVGSGESALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLSSRSQSRIVGSGESALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRQSPFMFYGAGYQVEKVIKSHPNYDSKTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSPFMFYGAGYQVEKVIKSHPNYDSKTKNDIALMKLQKPLTFNDL 360
QY 361 VKPVCLPDPGMMLOPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVCLPDPGMMLOPEQLCWISGWGATEEKGKTSSEVLNAAKVLLIETQRCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSGQSGPLVTSKNNIWMILIGDTSWGSCKAKAYRPGVGNMVF 480
Db 421 TPAMICAGFLQGNVDSGQSGPLVTSKNNIWMILIGDTSWGSCKAKAYRPGVGNMVF 480
QY 481 TDWIYQMRADG 492
Db 481 TDWIYQMRADG 492

RESULT 7
AAE18101
ID AAE18101 standard; protein; 492 AA.
XX
XX AAE18101;
XX
DT 07-MAY-2002 (first entry)
XX Human 20P1F12/TMPRSS2 mutant protein (R255Q).
DE Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCl; cell growth; neoplasm;
XX cancer; vaccine; human.
XX Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 255 /note= "Wild type Arg substituted with Gln"
FT
XX WO200204953-A2.
XX 17-JAN-2002.
XX 12-JUL-2001; 2001WO-US022168.
XX

PR 12-JUL-2000; 2000US-00615285.
XX (AGEN-) AGENSYS INC.
XX Saferran D, Raitano AB, Hubert RS, Jakobovits A, Faris M;
XX Challita-Eid PM;
XX WPI; 2002-154967/20.
XX
XX Examining a biological sample for evidence of dysregulated cellular
XX growth, comprises comparing the status of prostate-specific, androgen-
XX regulated, secreted serine protease, 20P1F12/TPRSS2, in a corresponding
XX normal sample.
XX
XX Example 10; Page; 161pp; English.
XX
XX The present invention relates to methods and compositions for the
XX diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
XX kidney cancer derived from or based on a normally prostate-specific,
XX androgen regulated, cell membrane associated secreted serine protease
XX termed 20P1F12/TPRSS2. The invention further relates to a method of
XX examining a biological sample for evidence of dysregulated cellular
XX growth comprises comparing the status of 20P1F12/TPRSS2 gene (also
XX designated 20P1F12-GTC1) in the sample to the status of 20P1F12/TPRSS2
XX in a corresponding normal sample. The invention also relates to 20P1F12/
XX TPRS2 polynucleotides and their corresponding proteins. Methods of the
XX invention are used for examining a sample such as blood, serum, stool,
XX urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
XX The dysregulated cell growth is indicative of bladder cancer, lung
XX cancer, kidney cancer or ovarian cancer. It is useful for identifying
XX evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
XX portion of 20P1F12/TPRSS2 are useful for inhibiting growth of a cell
XX expressing 20P1F12/TPRSS2 in a patient suffering from bladder cancer,
XX lung cancer, ovarian cancer or metastatic cancer. The present sequence is
XX human 20P1F12/TPRSS2 mutant protein (R255Q). Note: This sequence is not
XX shown in the specification but is derived from wild type 20P1F12/TPRSS2
XX protein shown as SEQ ID NO: 2 in figure 1 of the specification (AAE18096)
XX
SQ Sequence 492 AA;
Query Match 99.9%; Score 2713; DB 5; Length 492;
Best Local Similarity 99.8%; Pred. No. 1.4e-202;
Matches 491; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVEHPAQYIPSPVQYAPRVLTA 60
DB 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVEHPAQYIPSPVQYAPRVLTA 60
QY 61 SNPVVCTOPKPSGTVCTSKKALCIITLTGTLVGAALAGLLWFMGSKCSNSGIEC 120
DB 61 SNPVVCTOPKPSGTVCTSKKALCIITLTGTLVGAALAGLLWFMGSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDCVSHCPGGEENRCVRYGPNFLOVYSSQKSWHPVCQDDWNEY 180
DB 121 DSSGTCINPSNWCDCVSHCPGGEENRCVRYGPNFLOVYSSQKSWHPVCQDDWNEY 180
QY 181 GRAACRDGYKNFVSSGIVDDSGSTSFMLNTSAGNVDYKLYHSDACSSKAVVSLR 240
DB 181 GRAACRDGYKNFVSSGIVDDSGSTSFMLNTSAGNVDYKLYHSDACSSKAVVSLR 240
QY 241 CIACGVNLNSRQSRIVGSGSALPGAWPQVSLHVQNVHVCVGGSIITPEWITVAACVEX 300
DB 241 CIACGVNLNSRQSRIVGSGSALPGAWPQVSLHVQNVHVCVGGSIITPEWITVAACVEX 300
QY 301 PLANNPHTWTAAGILRQSFMFAGYQVEKVIHSHPNYDSKTKNDIALMKLQKPLTNDL 360
DB 301 PLANNPHTWTAAGILRQSFMFAGYQVEKVIHSHPNYDSKTKNDIALMKLQKPLTNDL 360
QY 361 VKPVLCPNPGMLOPEQLCWISGCGATEEKGKTSVLNAKVLIIETORCNSRYVVDNLI 420
DB 361 VKPVLCPNPGMLOPEQLCWISGCGATEEKGKTSVLNAKVLIIETORCNSRYVVDNLI 420
QY 421 TPAMTCAGFLQGNVDSGCGPLVTSKNNIWLIGTWSGSCAKAYRPGVYGNVWF 480

DB 421 TPAMTCAGFLQGNVDSGCGPLVTSKNNIWLIGTWSGSCAKAYRPGVYGNVWF 480
QY 481 TDWIYRQMRADG 492
DB 481 TDWIYRQMRADG 492
RESULT 8
ADB75591
ID ADB75591 standard; protein; 492 AA.
AC ADB75591;
XX 04-DEC-2003 (first entry)
XX Prostate cancer marker protein.
DE Prostate; cancer; cytostatic; gene therapy; marker.
KW Homo sapiens.
OS WO2003009814-A2.
XX 06-FEB-2003.
XX 25-JUL-2002; 2002WO-US023913.
XX 25-JUL-2001; 2001US-0307982P.
XX 22-AUG-2001; 2001US-0314356P.
XX 25-SEP-2001; 2001US-0325020P.
XX 12-DEC-2001; 2001US-0341748P.
XX 05-MAR-2002; 2002US-0362158P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Schlegel R, Monahan JE, Endege WO, Gamavarapu M, Gorbacheva B;
XX Hoerh S, Kamatkar S, Woney AM, Glatk X, Zhao X, Anderson D;
XX WPI; 2003-248033/24.
XX New nucleic acid molecule, useful for diagnosing or treating prostate
XX cancer.
XX Disclosure; SEQ ID NO 415; 99pp; English.
XX The invention relates to newly discovered cancer markers associated with
XX the cancerous state of prostate cells. Also disclosed is a method of
XX assessing whether a patient is afflicted with prostate cancer. The method
XX of the invention involves assessing whether a patient is afflicted with
XX prostate cancer by comparing the level of expression of a marker in a
XX patient sample and the normal level of expression of the marker in a
XX control non-prostate cancer sample, where a significant increase in the
XX level of expression of the marker in the patient sample and the normal
XX level indicates that the patient is afflicted with prostate cancer.
XX Nucleic acids of the invention are useful for diagnosing or treating
XX prostate cancer, and may be useful in gene therapy. Sequences given in
XX ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 492 AA;
Query Match 99.8%; Score 2711; DB 7; Length 492;
Best Local Similarity 99.6%; Pred. No. 2e-202;
Matches 490; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVEHPAQYIPSPVQYAPRVLTA 60
DB 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVTVVEHPAQYIPSPVQYAPRVLTA 60
QY 61 SNPVVCTOPKPSGTVCTSKKALCIITLTGTLVGAALAGLLWFMGSKCSNSGIEC 120

Db 61 SNPVCTQPKSPGTVCTSKKALCITLTLGTLFLVGAALAGLLWKFWSKCSNSGIEC 120
 Qy 121 DSSGTCINPNSWCDGVSHCPGGEDENRKCVLYGPNFILOVYSQKSWHPVQDDWNNY 180
 Db 121 DSSGTCINPNSWCDGVSHCPGGEDENRKCVLYGPNFILOVYSQKSWHPVQDDWNNY 180
 Qy 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVSLR 240
 Db 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACSSKAVSLR 240
 Qy 241 CIACGVNLNSRSQSRIVGGESALPGAWPMQVSLHVONVHVCGSIITPEWIVTAACHVEK 300
 Db 241 CIACGVNLNSRSQSRIVGGESALPGAWPMQVSLHVONVHVCGSIITPEWIVTAACHVEK 300
 Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
 Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
 Qy 361 VKPVCLEPNPQWMLQPEOLCWSGWGATEEKGKTSVLNAKVLIIETQCNRSRYVDNLI 420
 Db 361 VKPVCLEPNPQWMLQPEOLCWSGWGATEEKGKTSVLNAKVLIIETQCNRSRYVDNLI 420
 Qy 421 TPAMICAGFLQGNVDSQCGSDGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVMVF 480
 Db 421 TPAMICAGFLQGNVDSQCGSDGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVMVF 480
 Qy 481 TDWIYQMRADG 492
 Db 481 TDWIYQMRADG 492

RESULT 9

AAY92050
 ID AAY92050 standard; protein; 492 AA.

XX AC AAY92050;

XX DT 01-AUG-2000 (first entry)

XX DE HRPca6/7 polypeptide from androgen-inducible gene clone.

XX KW Androgen inducible; testosterone; prostate cancer; cytostatic; TMPRSS2; diagnosis.

XX OS Homo sapiens.

XX PN WO200018961-A2.

XX PD 06-APR-2000.

XX PF 30-SEP-1999; 99WO-US022535.

XX PR 30-SEP-1998; 98US-00163759.

XX PR 30-SEP-1998; 98US-00164159.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Macbeth KJ, Shyjan AW;

XX DR WPI; 2000-293182/25.

XX DR N-PSDB; AAA06803.

XX PT Novel methods for identifying compounds for treating prostate cancer comprising measuring the level of expression or activity of 1 or more of 11 genes or their products.

XX PS Claim 2; Fig 3; 108pp; English.

XX CC This protein is encoded by a gene which is androgen (e.g. testosterone) inducible in androgen-dependent prostate cancer cells (e.g. LNCap cells) and constitutively expressed in androgen-independent prostate cancer cells (e.g. LNC3 LNCap cells). Agents which decrease the expression or

CC activity of these clones may slow or arrest the growth of prostate cancer cells or may kill them. HRPca6/7 can be obtained from the sequence of the known gene for IMPRSS2. A compound useful for treating prostate cancer can be identified in a novel method comprising measuring the expression level, or activity, of HRPca2, 3, 6/7, 8, 9, 10, 13, 14, 15, 19, or peripheral-type benzodiazepine receptor (pBR) in a cell, in the presence and absence of a test compound. The sequences may also be used in diagnosis of prostate cancer and to determine efficacy of treatment for prostate cancer

XX SQ Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 3; Length 492;
 Best Local Similarity 99.8%; Pred. No. 2.8e-202;
 Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MALNSSGPPAIGPYIENHGVQENPYPAQPTVTVYVHPAQYVPSVPQVAPVLTQA 60
 Db 1 MALNSSGPPAIGPYIENHGVQENPYPAQPTVTVYVHPAQYVPSVPQVAPVLTQA 60
 Qy 61 SNPVCTQPKSPGTVCTSKKALCITLTLGTLFLVGAALAGLLWKFWSKCSNSGIEC 120
 Db 61 SNPVCTQPKSPGTVCTSKKALCITLTLGTLFLVGAALAGLLWKFWSKCSNSGIEC 120
 Qy 121 DSSGTCINPNSWCDGVSHCPGGEDENRKCVLYGPNFILOVYSQKSWHPVQDDWNNY 180
 Db 121 DSSGTCINPNSWCDGVSHCPGGEDENRKCVLYGPNFILOVYSQKSWHPVQDDWNNY 180
 Qy 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVSLR 240
 Db 181 GRAACRDMGYKNFYSSQGIIVDDSGSTSPMKLNTSAGNVDIYKLYHSDACSSKAVSLR 240
 Qy 241 CIACGVNLNSRSQSRIVGGESALPGAWPMQVSLHVONVHVCGSIITPEWIVTAACHVEK 300
 Db 241 CIACGVNLNSRSQSRIVGGESALPGAWPMQVSLHVONVHVCGSIITPEWIVTAACHVEK 300
 Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
 Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMKLQKPLTFNDL 360
 Qy 361 VKPVCLEPNPQWMLQPEOLCWSGWGATEEKGKTSVLNAKVLIIETQCNRSRYVDNLI 420
 Db 361 VKPVCLEPNPQWMLQPEOLCWSGWGATEEKGKTSVLNAKVLIIETQCNRSRYVDNLI 420
 Qy 421 TPAMICAGFLQGNVDSQCGSDGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVMVF 480
 Db 421 TPAMICAGFLQGNVDSQCGSDGGLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVMVF 480
 Qy 481 TDWIYQMRADG 492
 Db 481 TDWIYQMRADG 492

RESULT 10

AAM01315
 ID AAM01315 standard; protein; 492 AA.

XX AC AAM01315;

XX DT 04-OCT-2001 (first entry)

XX DE P1000C amino acid sequence.

XX KW Human; prostate cancer; prostate-specific; diagnosis; vaccine; cytostatic; gene therapy; metastasis.

XX OS Homo sapiens.

XX PN WO200151633-A2.

XX PD 13-JUL-2001.

XX PF 16-JAN-2001; 2001WO-US001574.

XX 14-JAN-2000; 2000US-00483672.
XX (CORI-) CORIXA CORP.
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
XX Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAM;
XX Wang A, Meagher MJ;
XX WPI; 2001-425873/45.
XX New polynucleotide encoding a prostate-specific protein, for diagnosing,
XX monitoring and treating prostate cancer in a patient and for use in
XX vaccines.
XX Claim 2; Page 539-540; 543pp; English.
XX The present invention describes polynucleotide sequences (I) which encode
XX prostate-specific proteins (II). (I) and (II) have cytostatic activity,
XX and can be used in vaccine production and gene therapy. (I), (II),
XX antibodies to (II), fusion proteins comprising (II), and isolated T cells
XX prepared using (I) or (II) are used treat cancer in a patient. (I) and
XX the antibodies are also used in the detection of cancer in a patient. The
XX cancer that is diagnosed or treated is particularly prostate cancer. (I)
XX and (II) can be used in vaccines. The antibodies or (I) can be used for
XX monitoring the progression of cancer in a patient. (I) and (II) can also
XX be used to improve diagnostic and therapeutic methods for prostate
XX cancer. They can indicate the level of metastasis as well as the prostate
XX volume. AAH93357 to AAH93944 and AAH01115 to AAH01318 represent
XX polynucleotide and amino acid sequences used in the exemplification of
XX the present invention
XX Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 4; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.8e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVYVHPAQYVSPVQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVYVHPAQYVSPVQYAPRVLTQA 60
Qy 61 SNPVCVCTQKPSGTCTKTKKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120
Db 61 SNPVCVCTQKPSGTCTKTKKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120
Qy 121 DSSGTCINPNCNCGVSHCPGEGEDENRCVLYGNSFVLQVYSSQKSWHPVCQDDWNEY 180
Db 121 DSSGTCINPNCNCGVSHCPGEGEDENRCVLYGNSFVLQVYSSQKSWHPVCQDDWNEY 180
Qy 181 GRAACRDMGYKNFYSSQGIYVDDSGTSFMKLNYSAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGIYVDDSGTSFMKLNYSAGNVDIYKLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLNSRQSRIVGSGESALPGANFQVSLHVQNVHVCVGSIIIPENIVTAACHVEK 300
Db 241 CIACGVNLNSRQSRIVGSGESALPGANFQVSLHVQNVHVCVGSIIIPENIVTAACHVEK 300
Qy 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVIHSPNDVSKTKNDIAIMKLOKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFYAGYQVEKVIHSPNDVSKTKNDIAIMKLOKPLTFNDL 360
Qy 361 VKPVCPLNPGMWLQPEQLCWISGWGATEBEKKTSEVLNAAKVLIIETQRNSRYVDNLI 420
Db 361 VKPVCPLNPGMWLQPEQLCWISGWGATEBEKKTSEVLNAAKVLIIETQRNSRYVDNLI 420
Qy 421 TPAMICAGFLOQNVDSOQDGGPLVTSKNINWILIGTWSGSCAKAYRPGVYGNVWF 480
Db 421 TPAMICAGFLOQNVDSOQDGGPLVTSKNINWILIGTWSGSCAKAYRPGVYGNVWF 480
Qy 481 TDWIYQNRADG 492
Db 481 TDWIYQNRADG 492

RESULT 11
AAU69960
ID AAU69960 standard; protein; 492 AA.
XX
XX AC AAU69960;
XX 30-JAN-2002 (first entry)
XX
XX DE Human prostate cDNA encoded protein #86.
XX Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX Homo sapiens.
XX OS
XX PN WO200173032-A2.
XX PD 04-OCT-2001.
XX PF 27-MAR-2001; 2001WO-US009919.
XX PR 27-MAR-2000; 2000US-00536857.
XX PR 09-MAY-2000; 2000US-00568100.
XX PR 12-MAY-2000; 2000US-00570737.
XX PR 13-JUN-2000; 2000US-00593793.
XX PR 27-JUN-2000; 2000US-00605783.
XX PR 09-AUG-2000; 2000US-00636215.
XX PR 29-AUG-2000; 2000US-00651236.
XX PR 06-SEP-2000; 2000US-00657279.
XX PR 02-OCT-2000; 2000US-00679426.
XX PR 10-OCT-2000; 2000US-00685156.
XX PR 09-NOV-2000; 2000US-00705729.
XX
XX (CORI-) CORIXA CORP.
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
XX Fanger GR, Retter MW, Stolk JA, Day CH, Vedwick TS, Carter D;
XX Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
XX WPI; 2001-639232/73.
XX N-PSDB; AAS64178, AAS64179, AAS64180.
XX New human prostate-specific polypeptides and polynucleotides useful for
XX the diagnosis and treatment of cancer, especially prostate cancer.
XX Claim 2; Page 573-574; 579pp; English.
XX The invention relates to isolated prostate-specific polynucleotides,
XX polypeptides, fusion proteins of the polypeptides, antibodies raised
XX against the polypeptides (or antigenic epitopes derived from them) and
XX antigen-presenting cells expressing the polypeptides. The antibodies are
XX useful for detecting the presence of cancer, especially prostate cancer.
XX The polypeptides, polynucleotides and the antigen-presenting cells are
XX useful for stimulating and/or expanding T cells specific for a tumour
XX protein, and for inhibiting the development of cancer especially prostate
XX cancer. Compositions comprising the polynucleotide and/or polypeptide are
XX useful for stimulating an immune response, and for treating cancer. The
XX oligonucleotide is useful for detecting cancer. The present sequence is a
XX prostate specific polypeptide of the invention
XX
SQ Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 4; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.8e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVYVHPAQYVSPVQYAPRVLTQA 60
Db 1 MALNSGSPPAIGPYENHGYQENPYPAQPTVTVYVHPAQYVSPVQYAPRVLTQA 60
Qy 61 SNPVCVCTQKPSGTCTKTKKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120

Db 61 SNPVCTQKPSGTVCTSKTKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
Qy 121 DSSGTCINPNSWCDGVSCHPCGGEDENRCVRLYGNFLOYTSSQKSHWHPVCOQDWNENY 180
Db 121 DSSGTCINPNSWCDGVSCHPCGGEDENRCVRLYGNFLOYTSSQKSHWHPVCOQDWNENY 180
Qy 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKXLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSFMKLNTSAGNVDIYKXLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLNSRSQSRIVGGESALPGAWPQVSLHVQNVHVCVGSIIIPWIVTAAHCVCK 300
Db 241 CIACGVNLNSRSQSRIVGGESALPGAWPQVSLHVQNVHVCVGSIIIPWIVTAAHCVCK 300
Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDYKNTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDYKNTKNDIALMKLQKPLTFNDL 360
Qy 361 VKPVLCPNPGMWLQPEQLCWISGWGATEBEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVLCPNPGMWLQPEQLCWISGWGATEBEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Qy 421 TPAMICAGFLQGNVDSQCGSDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWVF 480
Db 421 TPAMICAGFLQGNVDSQCGSDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWVF 480
Qy 481 TDWIYRQMRADG 492
Db 481 TDWIYRQMRADG 492
RESULT 12
ID ASU71851
XX ABU71851 standard; protein; 492 AA.
XX AC ABU71851;
XX XX
XX 10-JUN-2003 (first entry)
XX DE Prostate cancer associated protein #67.
XX KW Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
XX KW immunogen; cancer; prostate specific antigen; PSA;
XX KW prostatic acid phosphatase; PAP; prostate specific membrane antigen;
XX KW PSMA.
XX OS Homo sapiens.
XX PN US2002192763-A1.
XX PD 19-DEC-2002.
XX PF 29-JUN-2001; 2001US-00895793.
XX PR 04-OCT-1999; 99US-01574552.
XX PR 04-OCT-2000; 2000US-00679272.
XX PR 28-MAR-2001; 2001US-00822827.
XX XX
XX (XUJJ/) XU J.
XX (DILL/) DILLON D C.
XX (MITC/) MITCHAM J L.
XX (HARL/) HARLOCKER S L.
XX (JIAN/) JIANG Y.
XX (KALO/) KALOS M D.
XX (FANG/) FANGER G R.
XX (RETT/) RETTER M W.
XX (STOL/) STOLK J A.
XX (DAYC/) DAY C H.
XX (VEDV/) VEDVICK T S.
XX (CART/) CARTER D.
XX (LISX/) LI S X.
XX (WANG/) WANG A.
XX (SKEI/) SKEIKY Y A W.

PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
PA (HURA/) HURAL J. P. D.
PA (MCNE/) MCNEILL P. D.
PA (HOUG/) HOUGHTON R L.
PA (DBAS/) Y DE BASSOLS C V.
PA (FOYT/) FOY T M.
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Ronger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA, Hural J;
PI McNeill PD, Houghton RL, Y De Baseols CV, Foy TM,
XX WPI; 2001-245062/25.
XX DR Prostate specific protein and its encoding polynucleotide, useful for the
XX PT treatment and diagnosis of prostate cancer.
XX PS Example 2; SEQ ID NO 932; 85pp; English.
XX SS The invention describes a fusion protein comprising at least one amino
CC acid sequence of immunogenic portions of any of the 3 sequences not
CC defined in the specification, or sequences having at least 70 or 90 %
CC sequence identity to any one of the 35 sequences defined in the USPTO web
CC site, which is encoded by any of the 4 nucleotide sequences not defined
CC in the specification. The fusion protein, composition and methods are
CC useful for diagnosing, preventing and/or treating cancer, particularly
CC prostate cancer. The proteins are useful as markers to indicate the
CC presence or absence of cancer. This is the amino acid sequence of a
CC prostate cancer therapy associated protein. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20020192763
XX XX
SQ Sequence 492 AA;
Query Match 99.7%; Score 2709; DB 4; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.8e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MALNCGSPALPGYENHGYQENPYPAQTPVTVYEVHQAQYPSVPQAPRVLTA 60
Db 1 MALNCGSPALPGYENHGYQENPYPAQTPVTVYEVHQAQYPSVPQAPRVLTA 60
Qy 61 SNPVCTQKPSGTVCTSKTKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
Db 61 SNPVCTQKPSGTVCTSKTKALCITLTGLTFLVGAALAAAGLLWKFMSKCSNSGIEC 120
Qy 121 DSSGTCINPNSWCDGVSCHPCGGEDENRCVRLYGNFLOYTSSQKSHWHPVCOQDWNENY 180
Db 121 DSSGTCINPNSWCDGVSCHPCGGEDENRCVRLYGNFLOYTSSQKSHWHPVCOQDWNENY 180
Qy 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSEVLNAAKVLLIETQRCNSRYVDNLI 240
Db 181 GRAACRDMGYKNNFYSSQGIIVDDSGSTSEVLNAAKVLLIETQRCNSRYVDNLI 240
Qy 241 CIACGVNLNSRSQSRIVGGESALPGAWPQVSLHVQNVHVCVGSIIIPWIVTAAHCVCK 300
Db 241 CIACGVNLNSRSQSRIVGGESALPGAWPQVSLHVQNVHVCVGSIIIPWIVTAAHCVCK 300
Qy 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDYKNTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRQSFMYGAGYQVEKVISHPNVDYKNTKNDIALMKLQKPLTFNDL 360
Qy 361 VKPVLCPNPGMWLQPEQLCWISGWGATEBEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Db 361 VKPVLCPNPGMWLQPEQLCWISGWGATEBEKGTSEVLNAAKVLLIETQRCNSRYVDNLI 420
Qy 421 TPAMICAGFLQGNVDSQCGSDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWVF 480
Db 421 TPAMICAGFLQGNVDSQCGSDGGPLVTSKNNIWWLIGDTSWGSCKAKAYRPGVGNVWVF 480
Qy 481 TDWIYRQMRADG 492

Db	481 TDWIYQMRADG 492	PT	diseases, in particular prostate cancer, and as markers for the progression of cancer.
RESULT 13		XX	Claim 2; SEQ ID NO 932; 87pp; English.
ABB95420		XX	The present invention provides prostate-specific coding sequences and their encoded proteins. These can be used in the diagnosis and treatment of cancers, particularly prostate cancer. The present sequence is a protein described in the invention
XX	ABB95420 standard; protein; 492 AA.	XX	Sequence 492 AA;
AC	ABB95420;	XX	Query Match 99.7%; Score 2709; DB 5; Length 492;
DT	19-JUL-2002 (first entry)	XX	Best Local Similarity 99.8%; Pred. No. 2.8e-202;
XX	Human P1000C SEQ ID NO 932.	XX	Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
DE	Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant; gene therapy.	QY	1 MALNSGSPPAIGPYENHGYQENPYPAQPTVFTVEVHPAQYYPSPVQYAPRVLTOA 60
KW	Human sapiens.	Db	1 MALNSGSPPAIGPYENHGYQENPYPAQPTVFTVEVHPAQYYPSPVQYAPRVLTOA 60
OS	Homo sapiens.	QY	61 SNPVVCTOPKSPSGTVCTSKTKALCITLTGLTFLVGAALAGLLKPMGSKCSNGIEC 120
XX	US2002022248-A1.	Db	61 SNPVVCTOPKSPSGTVCTSKTKALCITLTGLTFLVGAALAGLLKPMGSKCSNGIEC 120
XX	21-FEB-2002.	QY	121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFVLQVYSSORCKSWHPVCDDWNNY 180
XX	12-JAN-2001; 2001US-00759143.	Db	121 DSSGTCINPSNWCDSVSHCPGDEENRCVLYGNFVLQVYSSORCKSWHPVCDDWNNY 180
XX	25-FEB-1997; 97US-00806099.	QY	181 GSAACRDMGYKNNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKKLYHSDACSSKAVVSLR 240
PR	01-AUG-1997; 97US-00904804.	Db	181 GSAACRDMGYKNNFYSSQGI VDDSGSTSFMKLNTSAGNVDIYKKLYHSDACSSKAVVSLR 240
PR	10-FEB-1998; 98US-00020956.	QY	241 CIACGVNLNSRQSRIVGSGALPGAWPQVSLHVQNVHVGSGSIITPEWITVTAHCVCK 300
PR	25-FEB-1998; 98US-00030607.	Db	241 CIACGVNLNSRQSRIVGSGALPGAWPQVSLHVQNVHVGSGSIITPEWITVTAHCVCK 300
PR	14-JUL-1998; 98US-00115453.	QY	301 PLNPNPWHMTAFAGILRQSFMYGAGYQVEKVISHPNYSKTKNNDIALMKLQKPLTFNDL 360
PR	23-SEP-1998; 98US-00159812.	Db	301 PLNPNPWHMTAFAGILRQSFMYGAGYQVEKVISHPNYSKTKNNDIALMKLQKPLTFNDL 360
PR	15-JAN-1999; 99US-00232149.	QY	361 VXPVCLPNPFGMMLQPEQLCWISGWGATEBKTSVINAAKVLLIETORCNSRYVYDNL 420
PR	09-APR-1999; 99US-00288946.	Db	361 VXPVCLPNPFGMMLQPEQLCWISGWGATEBKTSVINAAKVLLIETORCNSRYVYDNL 420
PR	13-JUL-1999; 99US-00352616.	QY	421 TPAMICAGFLQGNVDSQQDGGPLVTGKNNIWLLIGDTSWGGCGCAKAYRPGVYGNVMVF 480
PR	12-NOV-1999; 99US-00439313.	Db	421 TPAMICAGFLQGNVDSQQDGGPLVTGKNNIWLLIGDTSWGGCGCAKAYRPGVYGNVMVF 480
PR	18-NOV-1999; 99US-00443686.	QY	481 TDWIYQMRADG 492
PR	14-JAN-2000; 2000US-00483672.	Db	481 TDWIYQMRADG 492
PR	27-MAR-2000; 2000US-00536857.	RESULT 14	
PR	09-MAY-2000; 2000US-00568100.	ABR54532	ID ABR54532 standard; protein; 492 AA.
PR	12-MAY-2000; 2000US-00570737.	XX	AC ABR54532;
PR	17-JUN-2000; 2000US-00593793.	XX	DT 28-AUG-2003 (first entry)
PR	23-JUN-2000; 2000US-00605783.	XX	XX Prostate tumour specific protein SEQ ID 932.
PR	10-AUG-2000; 2000US-00636215.	XX	XX Cytostatic; gene therapy; prostate-specific protein; PSP; human;
PR	29-AUG-2000; 2000US-00851236.	XX	XX Immune response; prostate cancer.
PR	06-SEP-2000; 2000US-00857279.	XX	XX Homo sapiens.
PR	02-OCT-2000; 2000US-00679426.	XX	XX OS
PR	10-OCT-2000; 2000US-00685166.	XX	XX XX
XX	(XUJJ/) XU J.	XX	XX WO200289747-A2.
PA	(DILL/) DILLON D C.	XX	XX 14-NOV-2002.
PA	(MITC/) MITCHAM J L.	XX	XX 09-MAY-2002; 2002WO-US014753.
PA	(HARL/) HARLOCKER S L.	PF	
PA	(JIAN/) JIANG Y.		
PA	(KALO/) KALOS M D.		
PA	(FANG/) FANGER G R.		
PA	(RETT/) RETTER M W.		
PA	(STOL/) STOLK J A.		
PA	(DAYC/) DAY C H.		
PA	(VEDV/) VEDVICK T S.		
PA	(CART/) CARTER D.		
PA	(LISX/) LI S X.		
PA	(WANG/) WANG A.		
PA	(SKEI/) SKEIKY Y A W.		
PA	(HEPL/) HEPLER W T.		
PA	(HEND/) HENDERSON R A.		
XX	Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;		
PI	Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;		
PI	Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;		
XX	WPI; 2002-255649/30.		
DR	New prostate-specific polynucleotides for diagnosing and treating		
XX			
PT			

XX 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
XX (CORI-) CORIXA CORP.
XX Xu J, Dillion DC, Mitcham JL, Harlocker SL, Jiang Y, Henderson RA;
PI Kalos MD, Ronger GR, Retter M, Stolk JA, Day CH, Vedvick IS;
PI Carter D, Li SX, Wang A, Skeiky YAW, Hepler WT, Hural J;
PI McNeill PD, Houghton RL, Vinals Y De BassolsC, Foy TM, Watanabe Y;
PI Deng T;
XX WPI; 2003-167130/16.
XX New prostate-specific proteins and genes, useful in gene therapy,
PT particularly for stimulating an immune response in a patient, or treating
PT prostate cancer in a patient, as well as for diagnosing prostate cancer
PT in a patient.
XX Example 2; Page 642-643; 691pp; English.
XX The present invention relates to novel prostate-specific proteins (PSP)
CC and their coding sequences. The PSPs and their coding sequences are
CC useful for stimulating an immune response in a patient, or for treating
CC prostate cancer in a patient and for determining, detecting or diagnosing
CC the presence of a cancer in a patient. The present sequence was used to
CC illustrate the invention
XX
XX Sequence 492 AA;
SQ
Query Match 99.7%; Score 2709; DB 6; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.8e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MALNCGSPPAIGPVYENHGGQENPVPYPAQTVVTVYVHPAQYPSVPQYAPRVLTQA 60
Db 1 MALNCGSPPAIGPVYENHGGQENPVPYPAQTVVTVYVHPAQYPSVPQYAPRVLTQA 60
Qy 61 SNPVCTQPKSPSGTCTSKTKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120
Db 61 SNPVCTQPKSPSGTCTSKTKALCITLTGTLVGAALAAAGLLKFMGSKCSNGIEC 120
Qy 121 DSSGTCINPNCWCDGVSHCPGGEDENCVRLYGNFVLQVYSSQKSWHPVCCDDWNEY 180
Db 121 DSSGTCINPNCWCDGVSHCPGGEDENCVRLYGNFVLQVYSSQKSWHPVCCDDWNEY 180
Qy 181 GRAACRDMGYKNFYSSQGIYDSSGTSFMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
Db 181 GRAACRDMGYKNFYSSQGIYDSSGTSFMKLNSTAGNVDIYKLYHSDACSSKAVVSLR 240
Qy 241 CIACGVNLNSRQSRIVGGSALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAACHVEK 300
Db 241 CIACGVNLNSRQSRIVGGSALPGAMPQVSLHVQNVHVCVGGSIITPEWIVTAACHVEK 300
Qy 301 PLNPNWHTAPAGTLRQSFYAGYQVEKVIHSPNVDYKNDIALMKLQKPLTFNDL 360
Db 301 PLNPNWHTAPAGTLRQSFYAGYQVEKVIHSPNVDYKNDIALMKLQKPLTFNDL 360
Qy 361 VKPVCLENPNQMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLIIETQRCNSRYVDNLI 420
Db 361 VKPVCLENPNQMLQPEQLCWISGWGATEEKGKTSSEVLNAAKVLIIETQRCNSRYVDNLI 420
Qy 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWLIGDTSWGSCAKAYRPGVGNVWF 480
Db 421 TPAMICAGFLQGNVDSQCGSGPLVTSKNNIWLIGDTSWGSCAKAYRPGVGNVWF 480
Qy 481 TDWYRQMRADG 492
Db 481 TDWYRQMRADG 492

RESULT 15

ADB14382
ID ADB14382 standard; protein; 492 AA.
XX
AC ADB14382;
XX
DT 18-DEC-2003 (first entry)
XX Human prostate specific protein P1000C.
DE Human; prostate specific cDNA; cytostatic; immunostimulant; gene therapy;
KW cell therapy; vaccine; T-cell epitope;
KW class I major histocompatibility complex allele; MHC; prostate cancer;
KW tumour; antigen presenting cell.
XX
OS Homo sapiens.
XX
PN US2003185830-A1.
XX
PD 02-OCT-2003.
XX
PF 12-NOV-2002; 2002US-00294025.
XX
XX 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 09-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.
PR 09-APR-1999; 99US-00288946.
PR 13-JUL-1999; 99US-00352616.
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
PR 12-JAN-2001; 2001US-00759143.
PR 09-FEB-2001; 2001US-00780669.
PR 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
PR 09-MAY-2002; 2002US-00144678.
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Stolk JA, Kalos MD;
PI
XX WPI; 2003-756193/71.
XX N-PSDB; ADB14379, ADB14380, ADB14381.
XX New isolated polypeptide for use in a vaccine for stimulating an immune
PT response, or for treating or diagnosis cancer, preferably prostate
PT cancer.
XX
XX Example 2; Page; 101pp; English.
XX The invention relates to an isolated polypeptide comprising no more than
CC 11-542 amino acids of ADB13563 comprising a sequence ADB14487. The
CC peptides comprise a fragment ADB13563 of that contain naturally processed
CC T-cell epitopes for 3 class I major histocompatibility complex (MHC)
CC alleles. ADB13563 is a polypeptide encoded by a human prostate specific
CC cDNA, one of 648 disclosed as new. Also included are nucleic acids
CC encoding the proteins and peptides, expression vectors, a host cell
CC transformed with the vector, an isolated antibody (or antigen binding

CC fragment) that specifically binds to the protein or peptide, detecting
CC the presence of a cancer in a patient (comprising contacting a patient
CC sample with a binding agent that binds to the peptides or a polypeptide
CC appearing as ADB1358, detecting the amount of polypeptide that binds to
CC the agent and comparing the amount of polypeptide to a predetermined cut-
CC off value to determine the presence of cancer), a fusion protein
CC comprising the peptides or proteins, stimulating or expanding T cells
CC specific for a tumor protein comprising contacting T cells with the
CC peptides or the isolated T cell population, treating prostate cancer in a
CC patient comprising administering a composition comprising the peptides,
CC nucleic acids, antibodies or compounds, determining the presence of a
CC cancer in a patient and treating prostate cancer in a patient comprising
CC incubating cluster of differentiation (CD4+) and/or CD8+ T cells isolated
CC from a patient with the peptides or antigen presenting cells that express
CC (the peptides so that the T cells proliferate, and administering the
CC proliferated T cells to the patient. The peptides (or an oligonucleotide
CC that hybridizes to nucleic acid encoding them), is used to detect the
CC presence of cancer in a patient. The peptides, nucleic acids encoding, or
CC antigen-presenting cells expressing the nucleic acid, are used to
CC stimulate or expand T cells specific for a tumor protein. The peptides,
CC nucleic acids, antibodies, fusion proteins, T cell populations or antigen
CC presenting cells are used to stimulate an immune response or treat
CC prostate cancer in a patient. The present sequence is a prostate specific
CC protein of the invention. Note: Except where otherwise indicated, the
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docID=20030185830.
XX
SQ Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 7; Length 492;
Best Local Similarity 99.8%; Pred. No. 2 8e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALNSGPPAIGPYENHGYQENPYPAQPTVVPTVYEVHPAQYYPSPVQYAPRVLTQA 60
Db 1 MALNSGPPAIGPYENHGYQENPYPAQPTVVPTVYEVHPAQYYPSPVQYAPRVLTQA 60
QY 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLWKFWSKCSNSGIEC 120
Db 61 SNPVVCTQPKSPSGTCTSKTKALCITLTGLTFLVGAALAAAGLLWKFWSKCSNSGIEC 120
QY 121 DSSGTCINPSNWCDCVSHCPGEGEDENRCVRLYGSNFIQVYSSQKSWHPVCODDWNNY 180
Db 121 DSSGTCINPSNWCDCVSHCPGEGEDENRCVRLYGSNFIQVYSSQKSWHPVCODDWNNY 180
QY 181 GRAACRDMGYKNPYSSQGIYVDSGGSTSFMKLNTSAGNVDIYKLYHSDACSSKAYVSLR 240
Db 181 GRAACRDMGYKNPYSSQGIYVDSGGSTSFMKLNTSAGNVDIYKLYHSDACSSKAYVSLR 240
QY 241 CIACGVNLNSRQSRIVGGESALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
Db 241 CIACGVNLNSRQSRIVGGESALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAAHCVEK 300
QY 301 PLNNPWHWTAFAGILRSFMYGAGYQVEKVISHPNVDSTKNDIALMKLQKPLTFNDL 360
Db 301 PLNNPWHWTAFAGILRSFMYGAGYQVEKVISHPNVDSTKNDIALMKLQKPLTFNDL 360
QY 361 VKPCLNPNQMLQPEQICWISGKGATEEKGKTEVLNAAKVLIIETQCNRYVYDNL 420
Db 361 VKPCLNPNQMLQPEQICWISGKGATEEKGKTEVLNAAKVLIIETQCNRYVYDNL 420
QY 421 TPAMICAGFLOGNVDSCGDSGGLVTSKNNIWLIGDTSWGSCAKAYRPGVYGNMVF 480
Db 421 TPAMICAGFLOGNVDSCGDSGGLVTSKNNIWLIGDTSWGSCAKAYRPGVYGNMVF 480
QY 481 TDWYRQNRADG 492
Db 481 TDWYRQNRADG 492

Search completed: June 1, 2004, 14:34:32
Job time : 139 secs